Responses to the White House Office of Science and Technology Policy Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Federal Register :: Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

SUMMARY:

As described in the recent RFI on Clinical Research Infrastructure and Emergency Clinical Trials, the White House Office of Science and Technology Policy (OSTP), in partnership with the National Security Council (NSC), is leading efforts to ensure that coordinated and large-scale clinical trials can be efficiently carried out across a range of institutions and sites as needed to address outbreaks of disease and other emergencies. In this RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot, issued in partnership with the Office of the National Coordinator for Health Information Technology (ONC), OSTP and ONC seek input on viable technical strategies to distribute clinical trial protocols and capture clinical trial data using common application programming interfaces (APIs), in the pre-emergency phase as well as in emergency settings. One specific objective for this RFI is to gather information about whether there is value in a pilot or demonstration project to operationalize data capture in the near term, for example within 6-12 months of the close of comments on this RFI.

DATES:

Interested persons and organizations are invited to submit comments on or before 5:00 p.m. ET on December 27, 2022.

ADDRESSES:

Interested individuals and organizations should submit comments electronically to *datacollectionforclinicaltrials@ostp.eop.gov* and include "Data Collection for Clinical Trials RFI" in the subject line of the email. Due to time constraints, mailed paper submissions will not be accepted, and electronic submissions received after the deadline cannot be ensured to be incorporated or taken into consideration.

Instructions

Response to this RFI is voluntary. Each responding entity (individual or organization) is requested to submit only one response. Please feel free to respond to one or as many prompts as you choose.

Please be concise with your submissions, which must not exceed 10 pages in 12-point or larger font, with a page number on each page. Responses should include the name of the person(s) or organization(s) filing the comment.

OSTP invites input from all stakeholders including members of the public, representing all backgrounds and perspectives. In particular, OSTP is interested in input from health information technology (health IT) companies, app developers, clinical trial designers, and users of health IT products. *Please indicate which of these stakeholder types, or what other description, best fits you as a respondent.* If a comment is submitted on behalf of an organization, the individual respondent's role in the organization may also be provided on a voluntary basis.

Comments containing references, studies, research, and other empirical data that are not widely published should include copies or electronic links of the referenced materials. No business proprietary information, copyrighted information, or personally identifiable information should be submitted in response to this RFI. Please be aware that comments submitted in response to this RFI may be posted on OSTP's website or otherwise released publicly.

In accordance with FAR 15.202(3), responses to this notice are not offers and cannot be accepted by the Federal Government to form a binding contract. Additionally, those submitting responses are solely responsible for all expenses associated with response preparation.

FOR FURTHER INFORMATION CONTACT:

For additional information, please direct questions to Grail Sipes at 202-456-4444 or *datacollectionforclinicaltrials@ostp.eop.gov*.

SUPPLEMENTARY INFORMATION:

Background on emergency clinical trial research: OSTP (in partnership with the NSC and other Executive Office of the President components) is leading an initiative to enhance U.S. capacity to carry out clinical trials in emergency situations. This initiative is undertaken in accordance with the 2022 National Biodefense Strategy for Countering Biological Threats, Enhancing Pandemic Preparedness, and Achieving Global Health Security [1] and aligns with the goals of the American Pandemic Preparedness Plan (AP3).[2]

In the recent RFI on Clinical Research Infrastructure and Emergency Clinical Trials, OSTP is seeking input on the emergency clinical trials effort generally, including U.S.-level governance models to support the emergency clinical trials effort. Governance functions might include determining when coordinated, large-scale clinical research is needed, including research on countermeasures, to address outbreaks of disease or other biological incidents. A further governance function might be to develop clinical trial protocols (in coordination with external stakeholders), which could range from relatively simple studies to more complex ones involving the evaluation of investigational agents. OSTP also seeks comment in the RFI on Emergency Clinical Trials on how emergency clinical trial data should be managed to facilitate researchers'

access and analysis of results. One potential model would be the use of a centralized data repository and biorepository for specimens collected during trials.

In this RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot, to further prepare the U.S. clinical trials enterprise to carry out coordinated, potentially large-scale research protocols in an emergency setting, OSTP is seeking input on how best to operationalize protocol distribution and data capture from a technical perspective. Specifically, in this RFI we seek input on viable technical strategies to distribute clinical trial protocols and capture clinical trial data using common Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR®)-based APIs, in the pre-emergency phase as well as in an emergency setting. We seek comment on how to build towards both of these goals in a data capture pilot or demonstration project. This pilot, if implemented, could provide training for sites in underserved communities, thereby enlarging and strengthening the overall clinical trials infrastructure.

Desired use case: OSTP is still in the process of collecting information on governance models and other aspects of the emergency clinical trials initiative. For purposes of responding to this RFI, however, we would like responders to consider the following multi-step use case.

1. A U.S.-level governing entity would oversee development of a clinical trial protocol for broad distribution across clinical trial networks and sites.

2. Study sites would enroll participants in the trial (potentially using software mechanisms that can alert sites to potential subjects for a specific protocol in a manner that increases the diversity of trial populations). Sites would obtain appropriate e-consents and authorizations from participants.

3. Clinical trial data is typically sent to the trial sponsor though an electronic case report form (eCRF), which is the record of data that is required under the protocol to be captured for each trial participant. A data element in an eCRF is the smallest unit of observation for a particular subject.

4. The eCRFs would be transmitted electronically via common APIs to the sponsor.

5. The study site's health IT system would present the eCRF content to clinicians in a manner that expedites data collection and (ideally) fits within clinician workflows.

6. As the clinician obtains data elements to complete the eCRF, that data would be captured in the patient's electronic health record.

7. The clinical trial data would also be sent to a central data repository or small set of data repositories for researchers to analyze. It would be sent via common APIs so that researchers can easily interpret the eCRF data elements. Commercial cloud solutions are likely to house the data repository or repositories. Nonetheless, we would like a solution that would work across multiple cloud vendors.

For the purposes of this RFI, we are interested in the feasibility of all steps in the above hypothetical use case; we would also like input on how much of the use case could be operationalized in a pilot or demonstration project that might move forward in a timeframe of 6-12 months from the close of comments on this RFI.

ONC standards for interoperability: We believe that a pilot or demonstration project such as described above would be well supported by the regulatory and governance structure for interoperability of electronic health records (EHRs) that has been put in place by the Office of the National Coordinator for Health Information Technology (ONC). Among other initiatives, ONC is currently supporting development of the United States Core Data for Interoperability (USCDI) standard; the FHIR application programming interfaces (APIs); and Substitutable Medical Applications and Reusable Technologies (SMART) platform technologies that are compatible with FHIR interfaces and have given rise to a category of "SMART on FHIR" APIs. Certified health IT developers seeking certification on their Health IT Modules are currently working to meet various ONC certification criteria intended to improve data interoperability. For example, certified developers are required to implement certified API technology capable of patient and population services based on FHIR Release 4, the FHIR US Core Implementation Guide, and based on the HL7 FHIR® Bulk Data Access (Flat FHIR®) (v1.0.0: STU 1), August 22, 2019 Implementation Guide, by December 31, 2022.

In addition, ONC published the Trusted Exchange Framework, Common Agreement—Version 1, and QHIN Technical Framework—Version 1 on January 19, 2022. The overall goal of the Trusted Exchange Framework and Common Agreement (TEFCA) is to establish a universal floor for interoperability across the country. The Common Agreement will establish the infrastructure model and governing approach for users in different networks to securely share basic clinical information with each other—all under commonly agreed-to expectations and rules, and regardless of which network they happen to be in. Entities seeking to be designated as Qualified Health Information Networks (QHINs),[3] per the Common Agreement, can apply for that designation on a voluntary basis. A QHIN is a network of organizations that work together to share health information. The goal of TEFCA is for QHINs to connect directly to each other to ensure interoperability between the networks they represent and to serve a wide range of end users.

The Common Agreement defines Exchange Purpose(s) [4] as "the reason, as authorized by this Common Agreement including the Exchange Purposes SOP [5] , for a Request, Use, Disclosure, or Response transmitted via QHIN-to-QHIN exchange as one step in the transmission." Although research is not an authorized Exchange Purpose under the current version of the Common Agreement, it is a planned future Exchange Purpose, and responses to this RFI could inform how TEFCA might best support research in the future.

The implementation SOPs for Public Health and some other current Exchange Purposes, including Payment, Health Care Operations, and Government Benefits Determination, have not yet been developed. These SOPs will need to specify constraints, and at least some of the to-be-defined constraints are likely to be applicable to a future research-focused Exchange Purpose. Therefore, this RFI also seeks input on how TEFCA's Public Health Exchange Purpose

Implementation SOP might be designed to enable public health authorities to answer questions that align with the activities described in this RFI.

More information on ONC data interoperability initiatives is available at *https://www.healthIT.gov,* and more specific information about TEFCA at *https://www.healthit.gov/TEFCA* and *https://rce.sequoiaproject.org/.*

Information Requested: OSTP invites input from all interested parties as outlined in the instructions. Respondents may provide information for *one or as many topics* below as they choose.

Our goal for this RFI is to support optimized data collection for clinical trials carried out across a range of institutions and sites, both in emergency settings and in the pre-emergency phase, under the use case described above. We also seek input specifically on the value of designing a pilot or demonstration project to operationalize data capture in the near term, for example within 6-12 months of the close of comments on this RFI. With those goals in mind, we request input on the following topics:

1. United States Core Data for Interoperability (USCDI). We seek input on how U.S. Government and external stakeholders might leverage USCDI and future extensions of USCDI standards (such as USCDI+, an extension that supports federal partner program-specific requirements) to support emergency clinical trial research. It would also be helpful to receive comment on areas in which additional extensions might be necessary.

2. HL7 FHIR APIs. We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-emergency phase, and in what areas further advances might be needed. Specific topics in this connection include:

a. Use of an API that supports FHIR Bulk Data Access to support clinical research; whether bulk data exports from EHR systems can be used to support certain clinical trial protocols.

b. Use of the FHIR Questionnaire and QuestionnaireResponse resources to support clinical research.

3. SMART on FHIR APIs: We seek input on how U.S. Government and external stakeholders might leverage SMART on FHIR APIs, and in what areas further extensions might be needed. It would be helpful to receive comments on:

a. The most promising ways to create SMART on FHIR technologies that are portable across different institutions and EHR systems, but also provide adequate functionality to support emergency clinical trial research.

b. Whether the portability of SMART on FHIR tools provides a way to reach institutions and sites that have limited information technology resources; any promising ways to use SMART on FHIR to expand clinical research into underserved settings.

4. Clinical Decision Support (CDS) Hooks: We seek comments on how the HL7 CDS Hooks specification might be used to support clinical research, for example by creating prompts within the practitioner workflow during interaction with patients; and any advances that might be needed to support the use case described above.

5. Operationalizing protocols of varying complexity. As noted above, emergency clinical trial designs could range from relatively simple protocols to more complex studies involving the evaluation of investigational agents. We would appreciate comments on the following topics:

a. Whether any of the tools described above might be particularly well suited for certain types of studies.

b. For example,

i. Whether a bulk FHIR API export could be used to gather data for a simple trial protocol that is relatively close to the standard of care for a particular condition.

ii. Whether a FHIR Questionnaire/QuestionnaireResponse or a SMART on FHIR form would be useful in capturing data for a more complex protocol, such as one that involves an investigational agent.

c. Any technical limitations that we should be aware of regarding use of the above tools to operationalize clinical trial protocols.

6. Consent, deidentification, return of results. The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution.

a. In light of this, we seek comment on how the tools described above can be used to obtain, collect and/or manage any required informed consents and/or authorizations from patients or individuals in accordance with applicable regulations.

b. We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information. It would be helpful to receive comments on which deidentification and protection approaches are sufficiently mature to support a pilot effort in the near term.

c. Ideally, patient authorization would allow clinical trial data to be used for additional research beyond the original study. We would appreciate input on how the content collected for consent and authorization as well as the interfaces with deidentification technologies should be designed to enable flexible and responsible reuse of clinical trial data.

d. We seek comment on any technical capabilities that could support return of results to study sites or participants, where appropriate.

e. We seek comment on any regulatory or ethical guidelines that are relevant to patients' consents and authorizations under the use case described in this RFI, and on ways in which technical solutions might help ensure adherence to applicable regulatory or ethical guidelines.

7. User interface and experience. With all of the above technologies, we seek input on:

a. The best way to optimize the experience of health care providers, administrators, and other users, so as to maximize the utility and uptake of the product.

b. To the extent a particular form, app or other tool requires input from a health care provider or other user, the best ways to increase the likelihood that users will actually provide that input. It would be helpful to receive comments on methods that are available for completing empty fields after the fact, or otherwise managing any missing data.

c. For clinicians and health IT users: what existing tools, apps, or processes you have found most usable and why.

8. *Capturing data elements required for clinical trial protocols.*

a. We seek comment on the most promising technical approaches that would leverage common APIs to translate a particular clinical trial's data elements into data elements captured by user-facing tools (*e.g.,* FHIR Questionnaire feeding into a SMART on FHIR form or application).

b. If a tool such as a FHIR Questionnaire, FHIR QuestionnaireResponse, or SMART form or app is used to capture required data elements in this way, we seek comment on whether that creates an effective method for "pushing out" a research protocol to investigators and sites.

c. It would be helpful to receive comments on how best to ensure compliance with regulatory requirements for eCRFs when designing interfaces for data capture.

9. TEFCA and QHINs. As noted above, TEFCA is in the implementation phase at this time. In the future, the TEFCA QHINs are expected to support implementation of the FHIR APIs (see the ONC Recognized Coordinating Entity's January 2022 FHIR Roadmap for TEFCA Exchange [6]). We would appreciate comment on the opportunities and challenges regarding development of API implementations toward the use case described above, particularly given the current status of TEFCA and QHIN participation. Specific topics in this connection include the following:

a. Certain policy and/or technical constraints will need to be specified for currently authorized Exchange Purposes under the Common Agreement (*e.g.,* Public Health). We seek comment on which of these constraints will also be applicable to a future research-focused Exchange Purpose.

b. Opportunities that may exist for using the initially authorized Exchange Purposes to accomplish the use case described in this RFI.

c. How the Public Health Exchange Purpose could be used to advance the goals of this RFI; what aspects of the use case described above might fall within the scope of the Public Health Exchange Purpose.

d. How a future research-focused Exchange Purpose could be structured to advance the goals of this RFI.

e. Other opportunities or constraints related to TEFCA that should be considered with regard to this RFI.

10. Emerging technologies. We welcome comments on any future technological developments we should anticipate. Relevant technical developments include but are not limited to differential privacy; federated machine learning; other technologies referenced in the recent OSTP RFI related to privacy-enhancing technologies (PET) (*see* **Federal Register:** Request for Information on Advancing Privacy-Enhancing Technologies); and technologies outside of the PET space. Specific topics in this area include:

a. How future technologies might affect the use case and underlying assumptions laid out in this RFI.

b. How future technologies might change the nature of the software architecture, data architecture, or potential data collection solutions for clinical trials.

11. Pilot or demonstration project. We seek comment on how the U.S. Government can best work with external stakeholders and developers to develop a pilot or demonstration project that will operationalize clinical trial data capture and serve as a basis and model for data collection in the event of an emergency. This pilot or demonstration project could also potentially support clinical research in the pre-emergency phase. Specific topics include:

a. Whether data can be managed through a central repository or small set of central data repositories; options for cloud-based data storage.

b. Technical options that might hold promise in the short term to enable researchers from diverse locations to analyze the data collected from multiple clinical trial sites. We also seek comment on any additional options that should be considered in the long term.

c. Whether any parts of the pilot would be appropriately supported as

i. A demonstration project with commercial partnership.

ii. A public-private partnership.

iii. An agency-funded program.

12. Specific commercial capabilities. Commenters who are developing a technology or product that might be relevant to any of the topics set forth above are welcome to include a description of that product. Comments about a specific technology or product should be limited to three pages or less.

Dated: October 25, 2022.

Stacy Murphy, Operations Manager.

Footnotes

1. 2022 National Biodefense Strategy for Countering Biological Threats, Enhancing Pandemic Preparedness, and Achieving Global Health Security (October 2022), section 4.1.4.

Back to Citation

2. First Annual Report on Progress Towards Implementation of the American Pandemic Preparedness Plan (September 2022), at 22-23.

Back to Citation

3. The Common Agreement defines a QHIN as "to the extent permitted by applicable Standard Operating Procedure(s) (SOP(s)), a Health Information Network that is a U.S. Entity that has been Designated by the RCE and is a party to the Common Agreement countersigned by the RCE." *See* Common Agreement for Nationwide Health Information Interoperability Version 1, at 10, 6 (Jan. 2022), *https://www.healthit.gov/sites/default/files/page/2022-*.

Back to Citation

4. *See* Common Agreement for Nationwide Health Information Interoperability Version 1, at 6 (Jan. 2022), *https://www.healthit.gov/sites/default/files/page/2022-*.

Back to Citation

5. The current version of the TEFCA "Standard Operating Procedure: Exchange Purposes" specifies that authorized Exchange Purposes under the Common Agreement and that SOP are: Treatment, Payment, Health Care Operations, Public Health, Government Benefits Determination, and Individual Access Services.

Back to Citation

6. *https://rce.sequoiaproject.org/three-year-fhir-roadmap-for-tefca/*.

Back to Citation

[FR Doc. 2022-23489 Filed 10-27-22; 8:45 am]

BILLING CODE 3270-F1-P

Response		Starting page number
1	ZS Associates and IgniteData Ltd.	1
2	HIMSS Electronic Health Record Association	11
3	The Sequoia Project	19
4	Privacy Analytics	21
5	TransCelerate BioPharma Inc.	33
6	Scientific Knowledge Accelerator Foundation	39
7	Acoer	45
8	EnVeil	53
9	Vibrent Health	63
10	YonaLink	73
11	TERIDA	76
12	Suncoast RHIO	79
13	Vanderbilt University Medical Center/REDCAP project team	82
14	Medidata Solutions	92
15	AccendoWave	99
16	Oracle Corporation	101
17	Healthcare Leadership Council	121
18	Infectious Diseases Clinical Research Consortium	123
19	Faro Health	126
20	MITRE Corporation	132
21	Merative	143
22	DataCubed	151
23	Epic	155
24	Good Science Project	163
25	Castor	166
26	Palantir	176
27	VHL7 FHIR Accelerator Program	186
28	Datavant, Inc.	196
29	NHLBI Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies	201
30	Verily Life Sciences	204
31	Consortium for State and Regional Interoperability	206
32	Crescendo Health	212
33	Keyrus	217
34	HealthEx	227
35	Quantum Leap Healthcare Collaborative and OpenClinica, LLC	231
36	Thoughtworks	241
37	RTI International	248
38	Health Record Banking Alliance (HRBA)	259
39	Gregory Hess, Henrik Toft Sorensen, John Acquavella, Erzébet Horváth-Puhó, and Sean Redmond	267

December 20th, 2022

Response to RFI Document No. 2022-23489: Data Collection for Emergency Clinical Trials and Interoperability Pilot

To the Office of the Science and Technology Policy:

ZS Associates and IgniteData Ltd. are pleased to submit our response to the RFI Data Collection for Emergency Clinical Trials and Interoperability Pilot (Document No. 2022-23489). We appreciate the agency's commitment to ensure that coordinated and large-scale clinical trials can be conducted efficiently across range of institutions and sites to address outbreaks of disease and other emergencies.

ZS and IgniteData teams welcome the opportunity to comment on the RFI. As a global management consulting and technology firm focused on transforming global healthcare and beyond, ZS partners with sponsors and providers to discover and develop innovative medicines that improve patients' lives. IgniteData has been pioneering the next-generation solution for study data interconnectivity to help researchers collect study data fast, enabling vital treatments to reach patients sooner. IgniteData's Archer platform is the Virtual Research Assistant enabling researchers to easily transfer clinically validated from provider's EHR to sponsors' EDC.

ZS has been partnering with IgniteData since January 2022, to transform patient data automation for clinical trials (https://www.zs.com/about/newsroom/zs-invests-in-ignitedata-to-transform-clinical-research). Together, we share OSTP's commitment to develop viable technical strategies to distribute clinical trial protocols and capture clinical trial data using common application programming interfaces (APIs), in the pre-emergency phase as well as in emergency settings. We look forward to collaborating with the Agency to support the successful use of innovative technologies to advance the interoperability of clinical care and clinical research. Please contact Dr. Qin Ye and Mr. Dan Hydes with any questions regarding these comments.

Sincerely,

ZS Associates
Dr. Qin Ye
Two Bethesda Metro Center
7514 Woodmont Avenue
Bethesda, Maryland 20814 USA
Email: qin.ye@zs.com
Phone: +1240 425 4003

IgniteData
Mr. Dan Hydes
5a Frascati Way
Maidenhead
SL6 4UY
Email: dan.hydes@ignitedata.co.uk
Phone: +44 118 453 0403

1. **United States Core Data for Interoperability (USCDI)**

IgniteData & ZS Comments:

The USCDI standard covers a broad range of the key data domains and data points typically required by a sponsor throughout a clinical trial. In terms of the use case described, the comprehensive adoption of USCDI across all sites involved in a clinical trial prior to any such study commencing will support the ability to efficiently capture the study data in a unified manner across all the participating sites.

The universal adoption of USCDI and standardisation of coding will significantly expedite the transfer of clinical trial data from multiple sites to a single or set of data repositories, allowing researchers to analyse this data in the most efficient way possible.

Using a combination of an EHR2EDC solution such as IgniteData's Archer and HL7 FHIR APIs it's possible to map between the location of a trial data point in each site's EHR FHIR structure and its desired location in the eCRF within the sponsor EDC. A standardised approach to this mapping provides several key benefits:

1. A uniform data standards and coding approach across multiple sites and regions supports interoperability across healthcare institutions both within and outside of the remit of a clinical trial.
2. The more sites involved in the trial who utilise the same data standards and coding approach, the easier the process of exporting trial data to the sponsor will be.
3. A consistent data standard and coding approach minimises the risk of potential error and associated sponsor queries – a standardised approach means ultimately that there is mutual agreement between site and sponsor on data standards, coding and mapping which leaves no room for interpretation or ambiguity. Queries should therefore be significantly reduced, mitigating delays to the overall trial timeline.
4. A network-wide ontology of mappings between FHIR structure locations and corresponding EDC destinations can be built, supporting the reuse of mappings for particular sites and sponsors. The more studies that are run, the more comprehensive and intelligent these ontologies will become, meaning that the setup of each subsequent study for mapped sites and sponsors will be more efficient than the last. This is particularly vital when considering pre-emergency or emergency trials.

As mentioned, the USCDI standard covers a broad range of the data domains typically required by a sponsor for the purposes of a trial, however to achieve the use case this RFI is aiming to accomplish, it would be prudent to perform a wider validation exercise to ensure that the USCDI standard covers all of the data points required for a trial.

2. **HL7 FHIR APIs.**

IgniteData & ZS Comments:

The emergence of HL7 FHIR APIs has been a catalyst for change for health informatics. Although there are still several limitations to the art of the possible, it has acted as the catalyst for the creation of a new path of opportunity. FHIR APIs could be used in an emergency and pre-emergency phase today, but with limitation due to data availability and quality at hospitals in both the US and globally.

As an example, IgniteData has been using their Archer EHR-to-EDC software by connecting it to HL7 FHIR APIs at hospitals. A rate limiting factor that has been identified is something referred to as 'eSource-readiness', in other words: does a hospital have the correct technology, data availability/quality, the right knowledge/skill, process and willingness to ensure that what comes out of APIs is good quality?

Assuming one can connect as required (which IgniteData have evidence of with collaborations with Duke and MSKCC) the art of the possible is quite exciting. IgniteData have actively and routinely started to have major US health systems push data domains such as labs, medications, vitals, and demographics on a regular basis, rapidly and with ease.

FHIR Bulk data APIs are new and very exciting. Although they do not allow for 'bulk searching', once an application has had a patient loaded, they allow for an array of ideas to be applied. Ongoing extraction of data once set up and approved by the hospital is one example, but there are many more safety and monitoring use cases to be considered.

3. **SMART on FHIR APIs:**
   IgniteData & ZS Comments:

SMART on FHIR is not technically an API in its own right, SMART is the combination of HL7 FHIR APIs and an authentication methodology, in this case OAuth2. IgniteData has leveraged SMART technology to drive the scalability of Archer, so we will discuss this case study here.

Archer uses a cloud-based management and business rules engine to build out the data extract required for clinical trial eCRFs; this is a powerful asset if there is a way to easily connect it to the hospitals. To do this IgniteData made the decision to create a SMART on FHIR app which connects to major EHR vendors (e.g., Epic, Cerner). This SMART app micro-service within our environment allows us to significantly reduce the barrier to entry at a typical hospital because the application is reviewed, accredited, and supported by the major EHR vendors. As a result, the level of additional governance required by hospitals to deploy the application is significantly reduced. It is also technically easy and fast to install in the EHR environment (providing the hospital has access to the required skillset within their teams).

Archer provides extremely strong portability. The major challenge typically comes back to the current state of a hospital's data and what really comes out of their FHIR APIs. To combat this, IgniteData have been working on an eSource readiness deployment guide for sites, which details all the challenges we've come up against from different hospitals (e.g., how are medications managed and are they disseminated to HL7 FHIR APIs correctly) to enable each new, or lesser resourced, hospitals to engage faster and more effectively.

A final key discussion point is **why should a hospital bother?** Even with the advent of SMART apps and the ease with which they can be installed, hospitals need to know their efforts are not going to be wasted. With IgniteData's Archer platform, hospital sites want to know that the product is approved for use by top sponsors of clinical research, and for this reason we have created a champion programme where we are working with organisations such as Janssen, Bayer, Sanofi, and AstraZeneca to create a 'critical mass' of adoption. US

based emergency government programmes could also be added to this dynamic to drive a true paradigm shift, and this is as important as the technology itself.

4. **Clinical Decision Support (CDS) Hooks:**
   IgniteData & ZS Comments:

CDS hooks are an interesting concept and one we have explored as a business. For our main use case in our EHR-to-sponsor technology we haven't had the need to embed prompts into the clinical workflows of a hospital site. However, there are potential emerging use cases when used in combination with HL7 Bulk API interfaces which are of interest to IgniteData and ZS.

Together, IgniteData and ZS are exploring a concept which would allow patients in specific care pathways at site to be 'loaded' into an environment like Archer. Once a patient has been loaded into the system, we can then use the bulk APIs to test against protocol criteria such as inclusion and exclusion requirements for studies running at the site.

This approach has its limitations as it is not a big data or big search type of system. There are other angles, such as prompts to enrol patients into specific studies or a 'checker' to see if a patient in eligible for a trial currently loaded into the system. These concepts are not new and have been seen in EHR vendors own solutions such as Cerner PowerTrials, although the as the research market still screams for ways to identify and recruit patients it feels as though no one has solved this problem once and for all.

5. **Operationalizing protocols of varying complexity.**
   IgniteData & ZS Comments:

Using IgniteData's Archer as a technical example, the system has been designed for the exact complexities mentioned in the question – it focuses on regulatory light protocols in real-world evidence but predominantly focuses on regulatory heavy areas such as phase 1-3 clinical trials.

Where investigational agents are being introduced, regulatory compliance and trial execution becomes significantly more complex. In many emergency response situations, if we use Covid as an example, the introduction of an intervention is going to be a requirement to study a new vaccine and to get this vaccine to a point where a license can be given as quickly as is safely possible. This means all the technology used in the solution stack needs to be GxP compliant and have been evidenced through validation studies to accurately transfer data. Other areas such as audit logging and monitoring also need to be considered, as this enables both trust in where the data was sourced from and provides researchers with the tools to know that data from source has not changed after it has been entered into a data capture system.

When a trial design is less complex, for example a real-world observational study, and the protocol matches standard of care closely, technology like the bulk API becomes of great interest. IgniteData are currently running a research project to investigate the use of bulk APIs for this exact reason in its Archer product. The limitation of the bulk API is that it is not a big search/query tool, therefore the system needs to know exactly which patients it is

querying for. To do this the patient must first be enrolled into a central system – this is where SMART applications or CDS hooks can be used as part of standard practice in a patient's care pathway to 'enrol' them into systems such as Archer. Once enrolled, it is then possible for the system to utilise bulk APIs on an ongoing basis to perform tasks on patients.

Using FHIR questionnaires is an interesting discussion. It might be an effective way for data capture when an IMP is in play, but it is critical to not lose sight of two critical factors:

1. Complexity of EDC:
    a. EDC provides sponsors of research with a lot of additionalities when running a trial above and beyond simple data capture. This could include randomisation, monitoring/query management, GxP compliant auditing, unit management/conversion. FHIR questionnaires may gain some simplicity, whilst losing key management tools.
2. Hybrid design:
    a. Studies in most scenarios will have to run where the patients are, this might be a hospital site without the correct technology and/or a site with the right technology. Therefore, using FHIR based questionnaire may become an overcomplication – e.g., there is a requirement to use EDC as part of INP data capture, so if FHIR questionnaires are used 'as well' is an additional complexity rather than a help.

6. **Consent, deidentification, return of results.**

IgniteData & ZS Comments:

Moving data into a central repository for secondary use is certainly technically feasible. Once consent has been obtained, the destination of that data transfer could be to one or more places (e.g., EDC, central secondary uses repository). It's also viable to have ongoing extracts using the bulk APIs post consent and enrolment into a software application controlling this process.

Consent would need much greater and broader discussion than can be provided within this RFI response, but there are many ways it could be handled. In the case of IgniteData's Archer software, we typically work post-consent for a patient on an interventional study. This means the hospital consents the patient into the study and once consent is given the hospital enrols the patient into the study in the Archer platform. This may not be the case for post-marketing observational studies where patient consent is not provided and the data is being processed in line with local laws (as would be the case for a traditional late-phase study). Where consent is not being provided for certain study designs it's important to note that the data processing tool being used must be processing data on behalf of the hospital site, this is especially so in the EU with regards to GDPR.

7. **User interface and experience.**

IgniteData & ZS Comments:

A key factor in optimizing the experience of healthcare providers and maximising their use and uptake of the product is involving them in early-stage discussions around a new solution. Change management is a large and complex task, and it is important to ensure that solution users are engaged at the point at which the value proposition and benefits of any

new solution are being discussed to ensure they are on board with the strategy surrounding the solution.

IgniteData work closely with healthcare providers throughout the setup of a trial using Archer to understand their workflows, pressures, and requirements on a granular level. Taking this holistic approach to understanding the full situation at sites facilitates user engagement, as we focus a huge amount of effort on fostering an environment of collaboration between ourselves, sites, and sponsors.

The input required by user when using Archer is extremely intuitive, sites require minimal training in order to understand the use of the solution. In addition, using Archer closely mirrors the way site users would work without Archer, meaning the process is easy to learn and still saves the user a huge amount of time.

Whilst the process is very similar, test studies show that the use of Archer provides a time saving of approximately 96% when compared to manual transcription. This valuable time saving can be used on more valuable tasks such as patient care.

For missing data points or fields that can't be mapped using Archer, site users can revert to their original manual process and transcribe this directly into the sponsor EDC.

8. **Capturing data elements required for clinical trial protocols.**
   IgniteData & ZS Comments:

There are two separate methodologies which can be assessed as part of this question. Firstly, we can think about data which is already routinely collected inside the EHR as part of routine care. Secondly, we can think about ways in which to capture more data through user facing tools within the EHR e.g., FHIR questionnaires.
The issue with the latter is that we're moving a problem which already exists in common data capture tools to the EHR, whilst losing all the advanced secondary benefits of these system. Examples of these benefits might be the monitoring and query management features. If these features were not deemed important to the emergency response programme, then perhaps embedding forms in the EHR for additional data capture could be a good methodology.

The former point above seeks to focus on data which is going to be regularly available within the EHR, such as labs, medication vitals and demographics. In many data intensive clinical trials, these data domains account for 45-65% of all data capture in the clinical study (Figure 1). Where new data for the study needs to be created, study data management platforms such as EDC can then still be used to manage the data entry of a study.

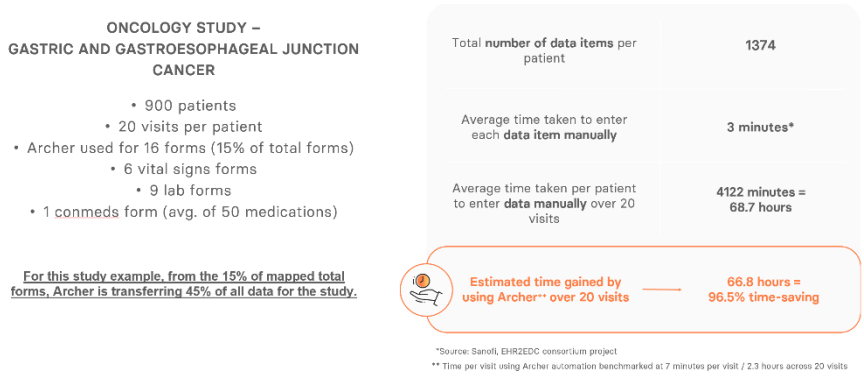


Figure 1, IgniteData projected results for Archer on AstraZeneca oncology study

We also need to consider that different hospital sites will have varying levels of technical capability and data availability. Therefore, we must assume that for many years to come the requirement for clinical studies in general using semi or fully autonomous data extraction will need a hybrid study design e.g., the final destination system for the data need to allow all sites to get the data in, regardless of technical state.

With software such as IgniteData's Archer product, if FHIR questionnaires are configured and used correctly at a hospital it is possible to call this data for the site to export to a third-party system. As above, the complexity around whether this is scalable solution in the real world today is something which need serious investigation against typical protocol requirements.

With regards to eCRF design for regulatory requirements, in our GxP compliant work with global biopharmaceutical organisations we are working directly with other GxP compliant data capture systems (e.g., Medidata Rave). To comment on how to keep forms within an EHR, as suggested in this question, is one of great challenge. This is because we begin to rely on large vendors, such as Epic, who are building a product which was not designed solely for the purpose of clinical research. There are ways in which one could envisage an embedded EDC type system for data capture using technology such as SMART and HL7 bulk APIs, but this would need careful consideration as to its scalability.

9. **TEFCA and QHINs.**

IgniteData & ZS Comments:

Whilst we understand that at present, according to Home - USQHIN there are no organizations who are yet qualified under the TEFCA model, we feel the ongoing qualification of organizations will play a vital role in the ability to pilot, validate and roll out a chosen technology solution.

The QHINs will likely provide a significant opportunity for market access for any chosen technology solutions and strategies, providing a platform which the OSTP can use both to expand solution use and uptake and to demonstrate the successes of the network. The anticipated support across the incoming QHINS of the implementation of FHIR APIs will put the network in a position of advantage when it comes to the rolling out of technology solutions which will support the interconnectivity of health data across QHIN sites.

Often one of the most complex and difficult elements of driving large scale change management programmes is related to building a critical mass of supportive stakeholders who help to drive adoption and uptake.

More automated transfer of regulatory-grade data from Electronic Health Records (EHRs) into study databases will accelerate the delivery of clinical trials and revolutionize the way clinical trials are conducted.

To support the uptake and adoption of this, IgniteData have formed an 'EHR2EDC Champion Programme' with a bold vision: to create a global network of connected champion hospitals within 3 years. The format for the Champion Programme is a pre-competitive collaboration and co-investment between 5 major pharmaceutical organisations, and will build on the foundations of multi-million USD investments by IgniteData since 2019.

The EHR2EDC Champion Programme offers the unique opportunity to refine the EHR-to-EDC/sponsor technology and processes that will help shape standards for the benefit of regulators and the wider industry for years to come.

10. **Emerging technologies.**

IgniteData & ZS Comments:

With regards to how future technologies might affect the use case and underlying assumptions laid out in this RFI, Differential Privacy is prone to losing its privacy guarantee when there is continuous data collection. There is a need for the development of novel algorithms/modifications to existing Differential Private solutions for sequential data.

Use-cases where adding noise to the data may limit the privacy guarantee (medical imaging) achieved by existing technologies. Further to this, multi-country trials may limit the usage of the model from one geography to another, hence implementing federated machine learning would not necessarily provide good accuracy and interoperability.

Good computation environments and the ease of distributed training may be key parameters in selecting sites/data nodes for participation in model training. Thus use-cases and trials would be required to adhere to these assumptions.

In relation to how future technologies might change the nature of the software architecture, data architecture, or potential data collection solutions for clinical trials, the following should be considered:

- Software Architecture
  - Able to run complex cryptographic algorithms for employing homomorphic encryption
  - Secure network architecture to share model parameters in a federated learning environment
  - Higher compute power as compared to other scenarios, since we will be developing local models based on data constraints as compared to a single compute where all the data is used for training models or analysing underlying patterns
  - Highly secure & private central server (for Federated Learning) for aggregation of models
- Data Architecture
  - Support usage of new technologies like Blockchain and Swarm Learning – maintenance of ledgers & smart contracts at the data nodes

- Since these technologies focus on building models at the different data nodes, it would require sufficient sample size at each data node which may be a challenge for sites with less patients, rare diseases, etc.

- Potential Data collection solutions
  - Pseudonymization can be used at the data nodes to increase privacy and preserving the underlying data properties, which will additionally lead to high utility
  - Synthetic data can help mitigate data access, availability & low sample size issues

11. **Pilot or demonstration project.**

IgniteData & ZS Comments:

IgniteData strongly believe that we have the tools, knowledge, and market access to help the U.S. Government develop a pilot or demonstration project. Many of the concepts discussed in this RFI are already conceptually part of a) the challenge we are already trying to solve; and b) already being demonstrated in real life in US and other global hospitals.

Through the pilot projects we have running with multiple global pharmaceutical companies and major health systems IgniteData have technology which is in place and being proven to a GxP standard. The technology is inherently designed to be scalable, adaptable and EHR and data capture system agnostic. This means we are not tied in to one EHR or one sponsor EDC. This provides the U.S government with a cost-effective adaptable suite of tools which can be rapidly developed to fit needs and 'plugged in' to the emergency response programme.

Using Archer, IgniteData and ZS would be able to connect to multiple hospitals and centrally process data through our cloud into which ever data capture system the U.S. Government deemed fit for the storage and further analysis of data.

12. **Specific commercial capabilities.**

IgniteData & ZS Comments:

IgniteData, who were founded in 2014, have developed Archer, a fully vendor agnostic, GxP compliant EHR2EDC solution which was developed following UK government funding in order to automate the structured data flow at hospitals for the purposes of clinical research.

Archer is a cloud-based Software-as-a Service ("SaaS") solution capable of remotely accessing and transforming patient medical records to enable the rapid, automated delivery of accurate, tracked, and validated data for delivery to multiple research applications. Regulatory-grade and clinically validated structured data from the site EHR systems will be mapped and pushed to the EDC system.

The value proposition for Archer is centred entirely around hospital sites, and ensuring that they are able to perform more studies, more efficiently, without the significant resource and administrative burdens that often cause unavoidable delays to the delivery of research.

Since its inception in 2019, Archer has gained significant traction with both hospital sites and pharmaceutical sponsors globally, including:

- 2020: Accredited by major EHR providers, including Epic and Cerner

- 2020: Topped the vendor rankings in the EHR2EDC Consortium vendor assessment led by AZ, Sanofi and Janssen
- 2021: Awarded a European pilot with AstraZeneca
- 2021: Successfully implemented Archer within several major UK NHS Trusts
- 2022: Multi-million dollar partnership with professional services company ZS, with the intent to power global Archer adoption
- 2022: Successfully implemented Archer at a major US research hospital
- 2022: Invited to partner with a major US cancer centre
- 2022: Ongoing programmes of work with an additional 5+ major pharmaceutical sponsors.
- 2022: IgniteData EHR2EDC Champion Programme launched

Archer acts as a virtual research assistant which allows research staff at hospital sites to export clinically validated, regulatory-grade data to the correct place in a pharmaceutical sponsor's EDC, without needing to manually re-enter the data.

The reduction in manual data entry massively reduces the delays that can occur between the point at which data is captured for a subject in a study and the point at which it is input into the sponsor EDC – enabling the sponsor to reach LVLP and database lock more quickly.

Sites can achieve efficiency gains of 96% whilst simultaneously improving data quality.

Reduced manual data entry means reduced risk of transcription errors, so sponsors are also able to reduce the amount of SDV and monitoring required to be performed throughout a clinical trial.

The platform is vendor agnostic and thanks to its advanced analytics engine, Archer can connect to any electronic system at a hospital which provides HL7® FHIR® based Application Programming Interfaces (APIs).

Security is integral to IgniteData and the Archer platform. The system is built around a 21 CFR Part 11 compliant audit database which tracks every action in the system, and critically, the provenance of the data transferred.

IgniteData is also an ISO27001, National Health Service (NHS) Data Security and Protection (DSP) Toolkit, General Data Protection Regulation (GDPR) and Cyber Essentials Plus compliant organisation, meaning we also comply with Health Insurance Portability and Accountability (HIPAA) requirements.

All data is transferred over an encrypted connection and no patient identifiable information can be viewed by IgniteData's application administrators - access is control is highly regulated, as per IgniteData's ISO27001 and NHS DSP Toolkit policies.

Finally, it is vital to note that Archer does not persist any patient identifiable data, the only data persisted is that required for audit logs – Archer is not a data lake or warehouse of any kind, and the sites retain full control of the export of their data to sponsors.

EHRA
HIMSS ELECTRONIC HEALTH RECORD ASSOCIATION

350 North Orleans Street, Suite S10000
Chicago, IL 60654

@EHRAssociation | ehra.org

December 21, 2022

The Honorable Dr. Arati Prabhakar
Director
Office of Science and Technology Policy
Executive Office of the President
Washington, DC 20500

Dear Director Prabhakar,

On behalf of the 30 member companies of the HIMSS Electronic Health Record (EHR) Association, we are pleased to offer our comments to the White House Office of Science and Technology Policy (OSTP) Request for Information (RFI) on *Data Collection for Emergency Clinical Trials and Interoperability Pilot.*

As a national trade association of EHR developers, Association member companies serve the vast majority of hospital, post-acute, specialty-specific, and ambulatory healthcare providers using EHRs and other health IT across the United States. Together, we work to improve the quality and efficiency of care through the adoption and use of innovative, interoperable, and secure health information technology.

While we acknowledge the great potential to utilize proposed tools and technologies to build an emergency clinical trial data collection infrastructure that could be used beyond emergency clinical trials, much effort is required to build the necessary implementation guidance and gain operational experience for rapid deployment. We urge OSTP to engage all critical stakeholders, particularly providers and their health IT suppliers, to address the complexities from the start.

The tools have promise but have not all been built for these use cases. We must not underestimate what it will take to establish a fully deployed infrastructure. An analogous effort around electronic prior authorization took two to three years to establish initial implementation guides, and initial implementations are only just starting for a limited scope of interactions, not yet the comparable full breadth of interactions. The experiences gained in those efforts can and should be taken advantage of, including other FHIR accelerator efforts, to optimize the reuse of common patterns and approaches.

We appreciate the opportunity to provide more detailed feedback as follows. The EHR Association and our individual members look forward to collaborating with you as this initiative unfolds.

Sincerely,

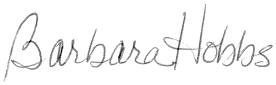
Hans J. Buitendijk
Chair, EHR Association
Cerner Corporation

David J. Bucciferro
Vice Chair, EHR Association
Foothold Technology

HIMSS EHR Association Executive Committee

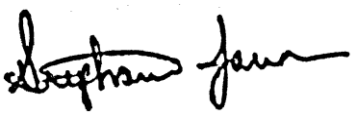
Pamela Chapman
Experity

William J. Hayes, M.D., M.B.A.
CPSI



Barbara Hobbs
MEDITECH, Inc.

Cherie Holmes-Henry
NextGen Healthcare



Stephanie Jamison
Greenway Health

Sasha TerMaat
Epic

Question 1: United States Core Data for Interoperability (USCDI)

The USCDI and USCDI+ extensions provide a useful framework to determine data for which FHIR-based support is available – or soon to be available, noting that when a USCDI version is published the actual FHIR-based implementation guidance necessary to support that USCDI version would not be available for another nine to twelve months. As USCDI versions, along with their supporting FHIR implementation guide standards, are included in either ONC's SVAP or certification rules, health IT developers focus on subsequent adoption and deployment. When included in updated certification rules, all certified health IT would aim to adopt that version, while they may or may not do so for versions referenced in SVAP. Therefore, the FHIR US Core version supporting the USCDI version referenced in certification rules would be the best indicator of data one can expect to be available through FHIR-based APIs in certified health IT once the adoption of that certification rule is mandated.

For example, as of January 1, 2023, it is reasonable to expect that all software certified to the 21st Century Cures Act Update to the 2015 Certification Rules (Cures Act Final Rule) supports the data required in FHIR US Core, at a minimum. For interoperability purposes, adherence to FHIR US Core is a more specific and relevant gauge than USCDI, as USCDI is only a set of concepts and vocabulary, not a standard on how to access and exchange that data. Only a standard such as FHIR or CDA C-CDA or v2 would provide that level of guidance.

The EHR Association suggests that the focus should be on identifying gaps in FHIR US Core to support the clinical trials for which uncurated data directly from the health IT source can be of value.

Question 2: HL7 FHIR APIs

The FHIR-based APIs being deployed in certified health IT, including individual data element and bulk data access, have the opportunity to support a wide range of data requests to inform clinical trials. FHIR-based APIs deployed for certification typically include QuestionnaireResponse as specified in FHIR US Core, even though USCDI v1, v2, or v3 do not include data using that resource. However, Questionnaire is not yet part of FHIR US Core, thus not as likely to be widely available across certified health IT. However, these tools would provide appropriate capabilities to access critical data more dynamically in support of clinical trials.

We should clarify that these APIs access the data as documented and do not distinguish between additional data or subsets of the data having been curated to ensure it is suited to clinical trials that have special data requirements. The data quality would be more aligned with what is suitable for real-world data-based research. The FHIR Questionnaire, QuestionnaireResponse, and Clinical Quality Language (CQL) would enable combining the ability to gather relevant data through automated processes where possible while allowing for further data collection through manual workflows. This could be facilitated through the source health IT or FHIR-based Apps that can orchestrate such automated and manual data collection using SMART Apps for user interactions as needed. HL7 FHIR Accelerators such as Vulcan (specifically focused on research and clinical trials), Da Vinci (focused on provider-payer interactions), as well as CDC's eCR Now, MedMorph, and NHSNlink initiatives demonstrate the direction and capabilities that can be pursued using FHIR-based technologies integrated into and/or connected with data sources that support FHIR US Core based APIs as a minimum.

FHIR US Core-based APIs are now widely deployed as part of certified HIT, while automated ingestion of FHIR Questionnaires and CQL translation into user interactions and automated data capture is starting to emerge, particularly among FHIR-based Apps.

Question 3: SMART on FHIR APIs

SMART on FHIR tools enable add-on solutions providing additional user-focused data collection for clinical trials where the source health IT may otherwise not (yet) collect such data, and support for these tools can connect to certified health IT. However, the source health IT would have to support both FHIR US Core and SMART to take advantage of those capabilities, thus still having some level of health IT capabilities. As referenced above, the type of App typically required would not solely be a SMART App but have other capabilities as well to orchestrate the clinical trial data requests.

Question 4: Clinical Decision Support (CDS) Hooks

CDS Hooks could be considered to streamline the initiation of data collection and sharing upon certain actions – including placing certain types of orders, documenting a qualifying condition, and other triggers that either potentially qualify the patient for a clinical trial or indicate the need for certain data collection for a patient within a clinical trial. It is critical to understand the workflows of interest in which such triggers occur and the type of interactions to consider depending on the variety of health IT that would be relevant. One cannot assume that all provider workflows and data are managed by a singular health IT solution, such as an EHR, as relevant data and triggers may be distributed across multiple systems.

Any FHIR-based implementation guides must clearly recognize the variety of health IT configurations that are reasonably expected to be deployed and thus needed to participate in the full workflow, starting with triggers and interactions relevant to the clinical trial at hand.

Initial CDS Hooks are starting to deploy across various health IT, although they are not addressed through certification criteria in the ONC's 21st Century Cures Update final rule.

Question 5: Operationalizing protocols of varying complexity

When considering FHIR-based tools and the types of studies for which they may be best suited, the key consideration may not be complexity, but volume. The challenge of the necessary data collection for a clinical trial often lies in the conditions of qualifying patients and specific data rather than the volume of data.

FHIR bulk data focuses on more efficient sharing of large data sets whether the data set was a result of simple data requirements or complex data requirements involving intricate conditions on qualifying data. The FHIR Questionnaire and CQL capabilities focus on the ability to convey simple data sets (FHIR Questionnaire) or more complex yet rigorously defined data sets (CQL).

As indicated in our response to Question 2, the automated ingestion of FHIR Questionnaire and CQL resulting in the automated collection of data through user interactions and/or FHIR API or native services are still emerging.

Question 6: Consent, deidentification, return of results

Where data needs to be shared in a de-identified format because sufficient authority and/or consent is not available to share identifiable data, the tools considered within this RFI can still be utilized. De-identification can start at the source, or in central/intermediary repositories that are authorized to manage identifiable data.

The primary challenge, however, is maintaining a complete patient record where clinical trials depend on aggregating data about the same patient across different source health IT, across different and distinct organizations that do not share a common enterprise master patient index (EMPI). Various techniques and technologies are available to utilize tokens or other privacy-preserving record linkages, but one must assess the risk of re-identification and how that risk can be managed.

Ensuring appropriate re-sharing/use of data for subsequent studies could be captured and conveyed using FHIR's Security Labeling capabilities. The challenges are not as much in the FHIR standards and sharing technologies, but rather in the upfront process of obtaining such patient consent and defining the scope and duration of such consent. Where data is being shared in de-identified form, any desired future changes to their consent would effectively be impossible,

whether expanding or contracting. This further emphasizes the need for great clarity and transparency when the patient is asked to consent to particular reuse of the data.

Where the data can be used in an identifiable form, such adjustments to its use could be managed, but the necessary standards and infrastructure to assert the most current patient consent directives relevant to their study would have to be established. One could consider the approaches being pursued by the San Diego LEAP project that is further advancing the use of patient-centric consent repositories that could incorporate consent relative to clinical trials as well, minimizing the places where a patient would have to maintain their various consent directives.

The ability to return data to study sites and participants could be enabled using FHIR-based technologies as well, including the emerging pub/sub capabilities that can be established at the time of joining a clinical trial.

Question 7: User interface and experience

It will be critical that any data collected for a clinical trial, particularly an emergency clinical trial when clinicians already are under great pressure, does not interfere or unduly add to a clinician's documentation burden. This will require significant consideration, as the EHR is the source of most clinical data.

Manual data collection must be minimized, if not eliminated, and should certainly not duplicate efforts when the data is readily available through automated means. Therefore, the clinical trial should be designed based on data already being documented, to the extent possible. This will enable maximum opportunities to automatically trigger the collection of relevant data and share it in identifiable, de-identifiable, or aggregate form.

We recognize that not all trials can rely on already available data and that clinicians are often willing and committed to performing the extra data collection. Well-defined use of FHIR-based tools has the opportunity to target the ideal users to collect the least amount of data, where the use of FHIR-based Apps (including SMART on FHIR Apps) can be made available with limited or no development efforts by the source health IT developer, assuming the health IT has minimally required FHIR based capabilities (particularly FHIR US Core based APIs and CDS Hooks).

To the extent that data remains identifiable, missing data could be collected and re-associated with the patient later.

Question 8: Capturing data elements required for clinical trial protocols

Considering the various analogous use cases that are starting to emerge, the anticipated flow would start with CDS Hooks invoking interactions with the research organization for the clinical trial at hand based on a patient cohort and/or defined patient characteristics. This is followed by a sharing of the FHIR Questionnaire identifying the relevant data of interest, either using specific questions to populate a form and/or CQL to specify the data of interest. Either the source health IT or a FHIR-based App will ingest that Questionnaire to then determine what data can be automatically gathered using individual FHIR US Core-based APIs or a bulk data export approach rather than requiring user interactions through a form of sorts.

Any workflow can be orchestrated to address missing data requiring follow-up, while then packaging data for sharing with the research organization. Depending on the extent to which the source health IT can translate FHIR Questionnaires or CQL into automated data collection or user interaction will determine the need for a FHIR-based App to be introduced to augment the source health IT.

Tools translating FHIR Questionnaire and CQL in automated data collection and/or user interactions are very much in the early days of development and utilization, although FHIR-based Apps in particular are starting to take advantage of this functionality utilizing FHIR-based APIs to collect the data automatically and interact with users using SMART on FHIR Apps for any data that otherwise could not be obtained.

We suggest that it is premature to consider regulatory requirements of FHIR-based capabilities at their current maturity level, including operational use for this type of use case. Rather, the availability of a comprehensive implementation guide is essential to start to progress a clear understanding of what is relevant and needed across all anticipated components of this infrastructure, including the multiple health IT present in the various provider organizations. Once sufficiently mature, with a clear understanding of the different roles that various health IT take on in the workflows can regulations, such as certification programs, effectively identify necessary and critical capabilities for applicable health IT, provider organizations, and research organizations.

Question 9: TEFCA and QHINs

TEF QHINs have the unique opportunity to identify where the patient has data and collect data across those locations. To the extent that clinical trials require access to a patient's data across multiple sources, TEF QHINs would provide a clear avenue to collect such data. Where the data of interest for a given patient is not distributed across different providers, TEF can still provide the legal and governance framework to ease connections with the provider of interest as well as using the clinical trial use case as one of the FHIR-based use cases in a TEF QHIN facilitated (but not brokered) FHIR-based interaction with the provider. This would reduce the number of

data-sharing agreements and interaction approaches to one agreement and one approach at a national level.

To the extent that the authority to access the data is under a public health authority, that Exchange Purpose could be used. This further emphasizes the need to align the techniques used in research and defined Exchange Purposes (including Public Health, Payment, Treatment, and Health Care Operations) to be consistent. Not all emergency clinical trials can or should be considered a Public Health Purpose, as explicit patient consent is required for participation where identifiable data is to be used. When unidentifiable or aggregated data is used, that may reduce these requirements, but raises privacy and ethical questions as to whether patients wish their data to be used in that manner beyond Treatment.

Question 11: Pilot or demonstration project

We suggest pilot activities should be explored in close collaboration with an HL7 FHIR Accelerator, such as Vulcan, to ensure continuous alignment in developing the necessary implementation guidance based on ongoing experience gained during connectathons and real-world pilots. It is critical that all relevant stakeholders have the opportunity to be engaged from the start and involve researchers, providers, and health IT suppliers at a minimum.

Question 12: Specific commercial capabilities

Given the nature of the EHR Association as a national trade association of EHR developers, we cannot directly respond to this question though individual member companies may provide further insights into their capabilities.

Generally, EHRs do provide a valuable source of relevant data when the data as documented can provide critical insights to clinical trials not requiring rigorous, trial-specific data collection. The use of FHIR-based apps connected to EHRs supporting FHIR and SMART can currently enable relevant incremental data collection. Advances are also being made such that EHRs over time could ingest data requests and potentially gather any additional data natively. The latter would vary by respective EHR developers as to the extent to which they do so or rely on FHIR-based apps to enable those capabilities more tightly integrated into their workflows.

Dr. Arati Prabhakar
Director of the Office of Science and Technology Policy (OSTP)
Eisenhower Executive Office Building
1650 Pennsylvania Ave NW
Washington, DC 20504
Notice Number: 2022-23489

December 27, 2022

Federal Register: Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Electronically submitted: datacollectionforclinicaltrials@ostp.eop.gov

Dear Director Prabhakar:

The Sequoia Project is pleased to submit comments to the Office of Science and Technology Policy (OSTP) regarding data collection for Emergency Clinical Trials and Interoperability Pilot.

The Sequoia Project is a non-profit, 501(c)(3) public-private collaborative that advances the interoperability of electronic health information for the public good. The Sequoia Project previously served as a corporate home for several independently governed health IT interoperability initiatives. We are also honored to have been selected by the Office of the National Coordinator for Health IT (ONC) to be the Recognized Coordinating Entity (RCE) for the Trusted Exchange Framework and Common Agreement (TEFCA). The comments and recommendations in this letter reflect this expertise independent of our role as the TEFCA RCE.

After careful review, we believe that additional development is needed to enable exchange for the purpose of data collection for emergency clinical trials within TEFCA. The use case described in this RFI does not fall within the scope of the currently authorized Public Health Exchange Purpose. The RCE and ONC would need to update the Common Agreement to incorporate an authorization-based or research focused Exchange Purpose, as well as develop an exchange purpose implementation standard operating procedure (SOP) to define other expectations and parameters for supporting this use case.

The RCE is working closely with ONC, other engaged agencies, and the private sector to develop standard operating procedures (SOPs) that would provide additional specifications for exchange related to the currently authorized Exchange Purposes. The RCE also plans to employ the same multistakeholder input process to bring additional Exchange Purposes into widespread use, including authorization-based use cases, on an expeditious, but deliberate pace.

The Sequoia Project is eager to collaborate with the OTSP on these topics and appreciates the opportunity to comment on how data collection for emergency clinical trials could fit into the TEFCA framework.

Sincerely,

Mariann Yeager

CEO, The Sequoia Project

WHITEPAPER

The SAFE Data Standard for Sharing Deidentified and Protected Clinical Trial Data

Pierre Chetelat, Research Associate
Luk Arbuckle, Chief Methodologist and Privacy Officer

We would like to thank the Office of Science and Technology Policy (OSTP) for this opportunity to respond to the RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot, issued in partnership with the Office of the National Coordinator for Health Information Technology (ONC). We believe that the pilot could lead to improved acquisition and distribution of clinical trial data and we appreciate that external input is being sought from external stakeholders, including members of the public.

Abstract

Since 2007, Privacy Analytics has been providing services and software in privacy-enhancing data sharing and analytics, including deidentification and protection approaches, for organizations in the consumer and healthcare industries. We are particularly interested in responding to topic 6b. of the RFI:

> *"We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information. It would be helpful to receive comments on which deidentification and protection approaches are sufficiently mature to support a pilot effort in the near term."*

In our response, we discuss a framework for standardizing the safe and responsible sharing of clinical trial data. The framework is founded on an approach to deidentification and protection which has been applied for decades and that has proven effective at protecting patient privacy.[1] For these reasons, this deidentification and protection approach could serve as the primary method for enabling the reuse of clinical trial data in the Pilot.

Below we present the SAFE Data Standard, a framework for protecting privacy to share non-identifiable clinical trial data for ethical secondary uses and disclosures. In our presentation, we draw heavily on an article that we contributed to about the standard in *Applied Clinical Trials*.[2] The SAFE Data Standard could help to preserve the utility of deidentified and protected data during the Pilot. Clinical trial data can be used and shared in different contexts, and these contexts determine in part the level of data transformation required as part of the deidentification process. For example, trial data may be made publicly available under an open data license on the European Medicines Agency (EMA) or Health Canada portal, or it may only be accessible to researchers on a platform such as Vivli, with strong screening and security controls. The more secure the context, the less the data needs to be transformed and the higher the data utility that can be preserved. The SAFE Data Standard makes it possible to consistently determine the appropriate level of data transformation given a specific context.

Deidentification and protection

Statistical (or quantitative risk-based) deidentification measures the probability of re-identifying individuals through indirectly-identifying pieces of information—such as demographic information, medical history, and medical event dates—and then reduces this probability through the use of various data transformations, such as shifting dates, generalizing disease classifications or demographic values, or removing (suppressing) outlier values in the data.[3]

The deidentification process renders data non-identifiable, such that the probability of re-identifying trial participants in the data is rendered very small.[4] A similar argument can be made for *anonymised* data under the GDPR (as originally outlined in the published article in Applied Clinical Trials). Identifiability can be viewed along a spectrum.[5,6] As the data are increasingly transformed, the identifiability of the data is gradually reduced until it reaches a level that is below the applicable deidentification threshold. At this point, the data are no longer identifiable. The appropriate threshold is determined based on data disclosure precedents, industry benchmarks, and/or regulatory guidance.

Many of the publicized re-identification attacks pertain to data that were minimally transformed or pseudonymized, with no other controls in place.[7-9] These examples demonstrate potential vulnerabilities and, as with any scientific discipline, serve as evidence to inform and evolve the field.[10] Statistical deidentification in consideration of all data variables and applicable technical and organizational controls is consistent with best practices and regulatory guidelines.

The level of identifiability in the trial data is determined by the similarity of participants in the data compared to the population. But contextual factors also matter. The more data a researcher or investigator has available to link or combine with the trial data, and the less restricted the use and environment of the trial data, the more likely re-identification becomes.

To support standardization, a process and framework for modelling data identifiability is needed to address a range of contextual re-identification opportunities. There are different ways in which identifiability can be modelled, and we opt to provide a conceptual representation of previously published and adopted statistical deidentification methodologies for measuring and managing re-identification risk. Industry consortia, such as PHUSE,[11] TransCelerate[12] and the Clinical Research Data Sharing Alliance (CRDSA),[13] play a role in promoting standardization in the exchanges of clinical trial data and may advance this conceptual representation to help meet practical implementation needs.

The process we adopt for measuring and managing identifiability is described by the equation Data x Context = Identifiability, where Data is the probability given a re-identification opportunity, and Context is the probability of a re-identification opportunity. This conditional probability establishes the level of identifiability of a data set in a particular context. Moreover, we can define an inequality based on an identifiability threshold that a data set must not exceed to be deemed deidentified and protected, using Data x Context < Threshold.

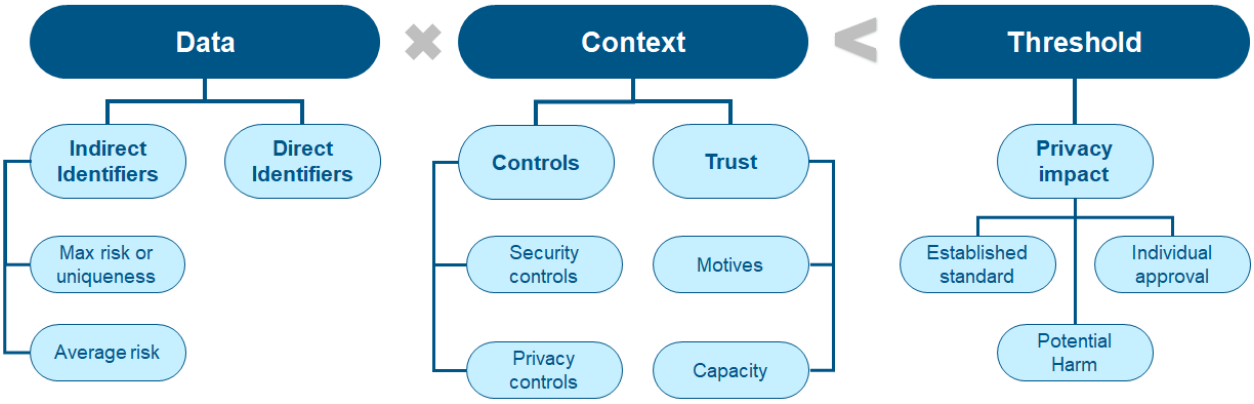


Figure 1 – Factors evaluated to deidentify and protect data within the applicable and appropriate threshold, where certain factors (eg, motives and capacity) are influenced by contractual controls and training obligations.

The conceptual framework and basis for standardization we introduce for the SAFE Data Standard can be extended to other forms of data, such as the outputs from remote query systems or synthetic data, assuming that privacy metrics can be established and enforced under varying contexts. As an example, Stadler, Oprisanu and Troncoso recently evaluated the use of differential privacy and found that the principles for assessing synthetic data are similar to those followed when assessing transformation methods for deidentification and protection. The authors thus demonstrate empirically that synthetic data does not provide a better trade-off between privacy and utility than transformation techniques to deidentify and protect data.[14] The practical goal in all cases is to identify the disclosure contexts shared frequently across clinical trial sponsors and align privacy metrics to the contextual risks associated with each, for consistency and greater standardization in how data are shared across these contexts.

Standardizing deidentification and protection levels

Because contextual factors—such as platform security and enforceable terms of use—influence the likelihood of re-identification, the degree to which data are transformed in the deidentification and protection process should be commensurate with these controls. However, without a common standard to define the degree of transformation, sponsors and platforms may adopt inconsistent methods, potentially resulting in unnecessary erosion of data utility or weaker privacy protection than needed.

To promote standardization and efficiency in the sharing of data, this paper proposed a SAFE Data Standard rating corresponding to a certain level of data transformation that can be used to quickly align stakeholders and effectively protect privacy. Because the design of a data-sharing portal (eg, security controls) and terms of use remain relatively constant over time for a single data platform, and certain characteristics of clinical trial data are constant, the level of data transformation needed to protect privacy can be standardized along a common scale from 0 to 5, where 0 is the raw trial data (often referred to as "coded" data due to clinical trials being blinded) and 5 is data transformed to the full extent required for access under an open data license or similar terms of use (eg, publication on EMA[15] or Health Canada transparency portals).

This concept is illustrated in Figure 2, with each rating having a defined context and degree of data transformation described further in the following sections. While data utility remains higher with statistical deidentification than with traditional methods such as redaction, the relative decrease in

data utility reflects the degree to which data are transformed to compensate for an absence of other mitigations (such as security controls).

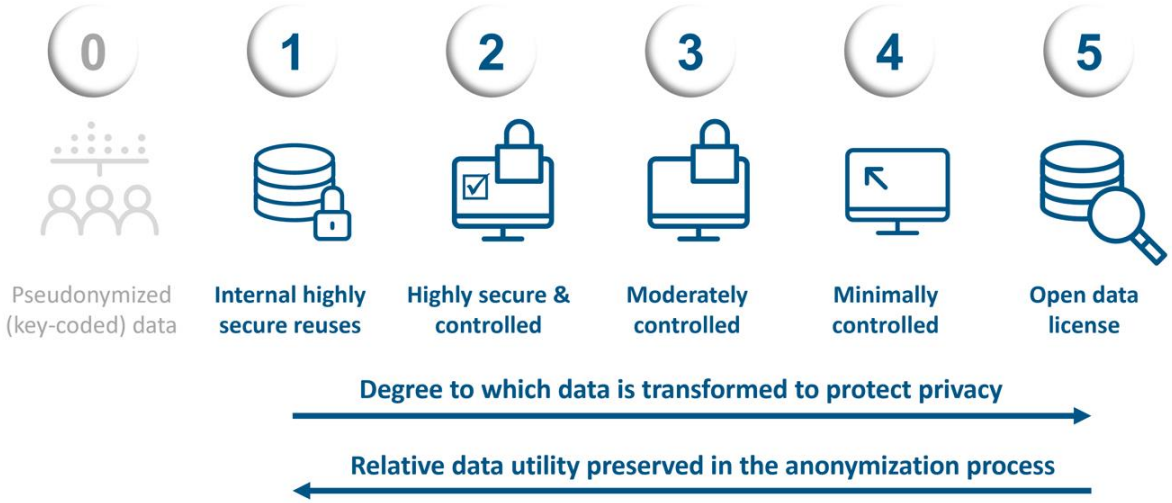


Figure 2 – Description of the SAFE Data Standard rating system, where data rated from 1 to 5 has been deidentified to reflect the context of data disclosure, with an increasing degree of data transformation and associated impact on data utility.

To maintain an adequate level of privacy, each level of data transformation on the 5-point scale also requires appropriate data protection measures, such as security and privacy controls and user contracts. The less the data are transformed, the greater the protections they will require. For each level on the scale, the standard specifies the appropriate measures for protecting the privacy of participants in the data. (Each of these levels is further defined in Figure 6 in the final section.)

If the data were made public without terms of use (eg, posted publicly on Google with no published terms of use), the data would have even less contextual protection than what is specified by a level 5 rating. The complete public release scenario is not addressed by the SAFE Data Standard, though it may warrant transformations greater than those recommended for level 5. If those accessing data do not agree to any terms of use, the data become even more susceptible to demonstration attacks. Demonstration attacks are typically launched by the media or academics striving to prove that re-identification is possible.[16,17] Given that an equivalent level of transparency can be attained through approaches adopted by the EMA and Health Canada, publishing clinical trial data with no terms of use is not typically required or recommended.

The 5-point rating is valuable because it communicates to all viewers not only the level of transformation to the data set itself, but also the protection measures that one would expect to find on a platform with a given rating. Moreover, the standard specifies the protection measures that platforms must implement to accommodate data at a particular utility level while maintaining adequate privacy. The result is a simplified concept of a numeric rating that can quickly be used to classify a data sharing platform.

While the primary focus of this SAFE Data Standard is on structured individual participant data given its analysis-friendly format, the term “data” and the use of the SAFE Data Standard can apply

more broadly to other information collected or produced during a clinical trial, including clinical study documents.

Data transformation

Each rating level prescribes an appropriate level of data transformation along two dimensions. The first dimension is the strict data tolerance, or equivalent minimum cluster size of similarly looking individuals across all trial participants. This is also known as a group size, which is related to the concept of equivalence classes in k-anonymity while accommodating different implementations of the same concept in complex data types, such as longitudinal clinical data.[18] The second dimension is the average data tolerance, or equivalent average cluster-size value, across all trial participants.

Cluster size is determined by the number of individuals who share the same indirectly identifying information. Figure 3 provides an illustrative example. In Figure 3, the highlighted data subjects form a cluster size of three since they all share identical values for the indirect identifiers of gender and year of birth.

	← Indirect identifiers →		← Non-identifying information →	
ID	**Gender**	**Year of Birth**	**Lab test**	**Lab result**
1	M	1959	Albumin, Serum	4.8
2	M	1969	Creatine Kinase	86
3	F	1955	Alkaline Phosphatase	66
4	M	1959	Bilirubin	<0
5	F	1942	BUN/Creatinine Ratio	17
6	F	1975	Calcium, Serum	9.2
7	F	1966	Free Thyroxine Index	2.7
8	F	1987	Globulin, Total	3.5
9	M	1959	B-type Natriuretic peptide	134
10	M	1967	Creatine Kinase	80

Figure 3 – Illustration of the cluster size concept, where the cluster size shown across two indirectly-identifying fields (gender and year of birth) is three due to three individual data subjects falling into this group.

If the minimum cluster-size value in this data set is two (strict tolerance level of 0.5), then every subject in the data set must have the same indirect identifier values as at least one other subject. In contrast, if the average cluster-size value is five (average tolerance level of 0.2), then the individuals in the data set must on average have the exact same indirect identifier values as four other subjects in the data. If a data set does not meet the desired minimum and average tolerance levels, then the indirect identifiers in the data set must be further transformed.

Average tolerances are relevant for private data-sharing releases in which the target of an adversary attempting re-identification could be any data subject (for example, an acquaintance such as an ex-spouse). The reason a strict condition is still applied to private releases is to ensure that no individual in the data is unique in the defined population.[18] The strict condition helps prevent "singling out" and is applied in private releases to indirect identifiers that may be used to single individuals out (eg, demographics). HHS Guidance Regarding Methods of Deidentification explicitly mentions the need to "determine the extent to which the subject's data can be distinguished in the health information."[4] Singling out would therefore be one such method that would always seem reasonably likely (if not a prerequisite) for the purposes of identification.[19-21]

Eliminating the ability to single out individuals is therefore a minimum condition, and, depending on the context and risks, average tolerance can then be evaluated for larger cluster sizes.

Tolerances also need to reflect real-world risks, which means evaluating cluster sizes for the population of individuals that gave rise to the data itself. The population we are concerned with is the one that contributes to the ability of someone to identify an individual in the shared or released data set. This may include the trial population, the population of similar trials, and the population in the same geographic area.[22] Cluster sizes to determine identifiability are therefore evaluated using statistical estimators for the defined population.

When data are being made public, the minimum cluster size is more applicable in the statistical modeling because demonstration attacks are a risk. In a demonstration attack, an individual's motive is to simply demonstrate that re-identification is possible, so the most identifiable record in the data is at greatest risk. Accordingly, for public releases, assume that an attack will occur and ensure a large minimum cluster size to protect against all types of attacks.

The ratings framework

A simplified rating system can standardize data utility and privacy for more effective, efficient, and trustworthy transparency. As presented, the data transformation rating specifies how much the data at each rating level have been transformed and gives an indication of how much utility the data have retained.

To ensure that a data transformation rating is appropriate for a given situation, the proposed assessment framework not only provides a rating scale for data transformations but also prescribes appropriate uses and contextual measures to accompany each data transformation level.

There are three distinct variables in the assessment framework that must be considered to determine whether a data transformation is situationally appropriate.

1. **Use:** The desired threshold is determined by the intended data use. Whether the data are being disclosed to external party(ies) or reused by the sponsor of the trial for new or secondary research can influence which threshold would be considered appropriate. There is established consensus on a threshold of 0.09 (cluster size of eleven) for external disclosures of trial data subject to basic terms of use. For internal reuses of data by a sponsor for further R&D beyond the original trial for which data was collected and consented to, a higher threshold can be argued for based on the extended participant and societal benefits of R&D and trial participants' general expectations of how sponsors use collected data.[23-28] Therefore, where the use (or reuse) of trial data by a sponsor is generally aligned to the purposes of collection from participants (eg, R&D), a higher threshold can preserve more utility. In all cases, appropriate justification should be documented by the sponsor along with the details of the implemented deidentification and protection strategy.

2. **Controls:** The data release context is defined in part by the extent of security and privacy controls in place as part of the data release. These controls prevent deliberate re-identification attempts and reduce the risk of a breach.

3. **Recipient Trust:** An assessment of the data release context also requires that one consider the degree to which recipients of data are known, trusted, and subject to enforceable terms of use that deter or prevent actions that would increase the likelihood of re-identification.

Enforceability refers to the sponsor's ability to impose financial and/or reputational consequences, through legal action, loss of funding, or otherwise, for non-compliance (eg, through a legally enforceable data sharing contract).

To standardize and achieve a common, defensible rating, these contextual factors need to be evaluated consistently. The following sections provide an assessment framework for controls and recipient trust.

If no controls are in place (for example, if data are being made available to the public for download), only the data need to be evaluated. However, for platforms or internal environments that do enforce privacy and data security, the following scale can be used to characterize the level of control. If the minimal level is not achieved (ie, if the basic controls are not in place), then a "zero control" context is assumed in determining the SAFE Data Standard rating.

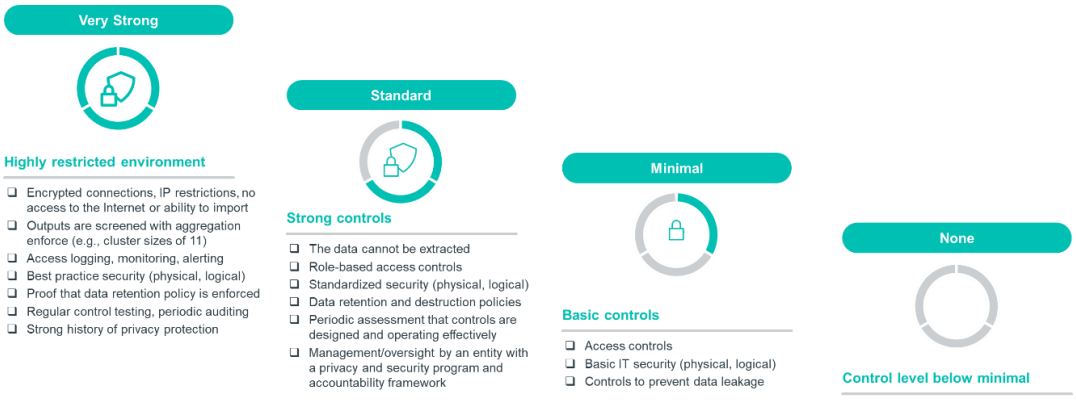


Figure 4 – Scale for evaluating levels of privacy and data security controls.

If there are no enforceable terms of use established with data recipients, nothing needs assessing. However, if data access is restricted to known entities who agree to terms of use, the following scale can be used to characterize the level of recipient trust. If the enforceable criteria are not demonstrated, then no recipient trust is assumed in the SAFE Data Standard rating.

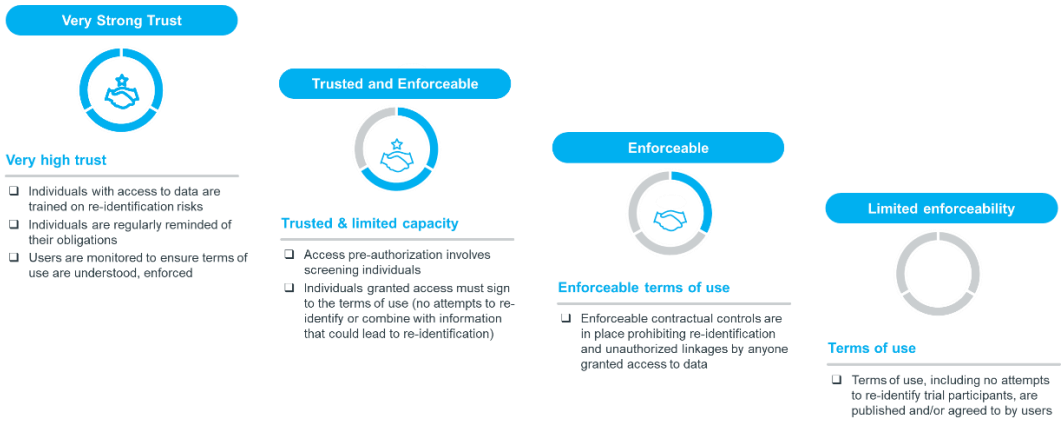


Figure 5 – Scale for evaluating levels of recipient trust.

In summary, the proposed data transformation rating is from 0 to 5, where 0 is the raw data and the scale from 1 to 5 reflects varying degrees of data transformation proportional to the type of use and contextual controls in place to protect data from re-identification opportunities. Consistently evaluating the uses of deidentified data and the protection context can speed and standardize deidentification and protection processes applied across sponsors and enforce a common baseline for privacy while maximizing data utility and analytic benefits. Table 2 summarizes the data transformation ratings from 0 to 5.

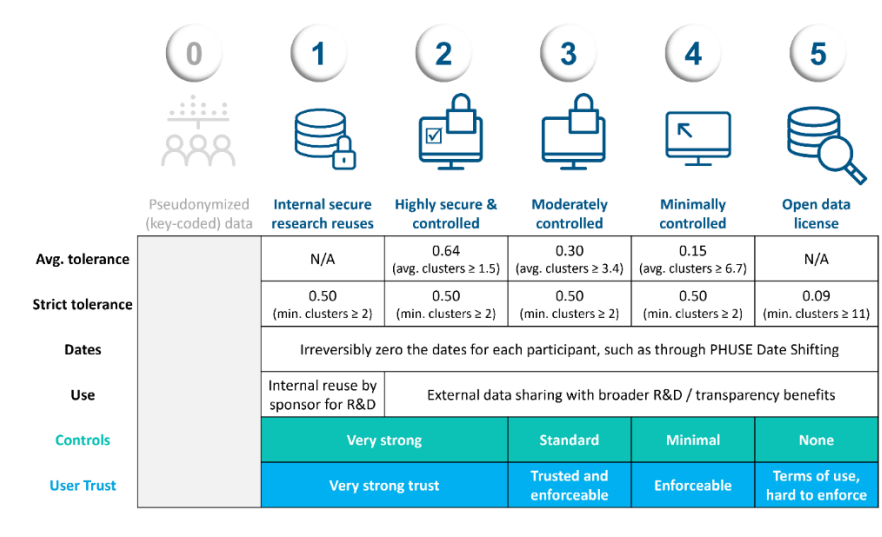
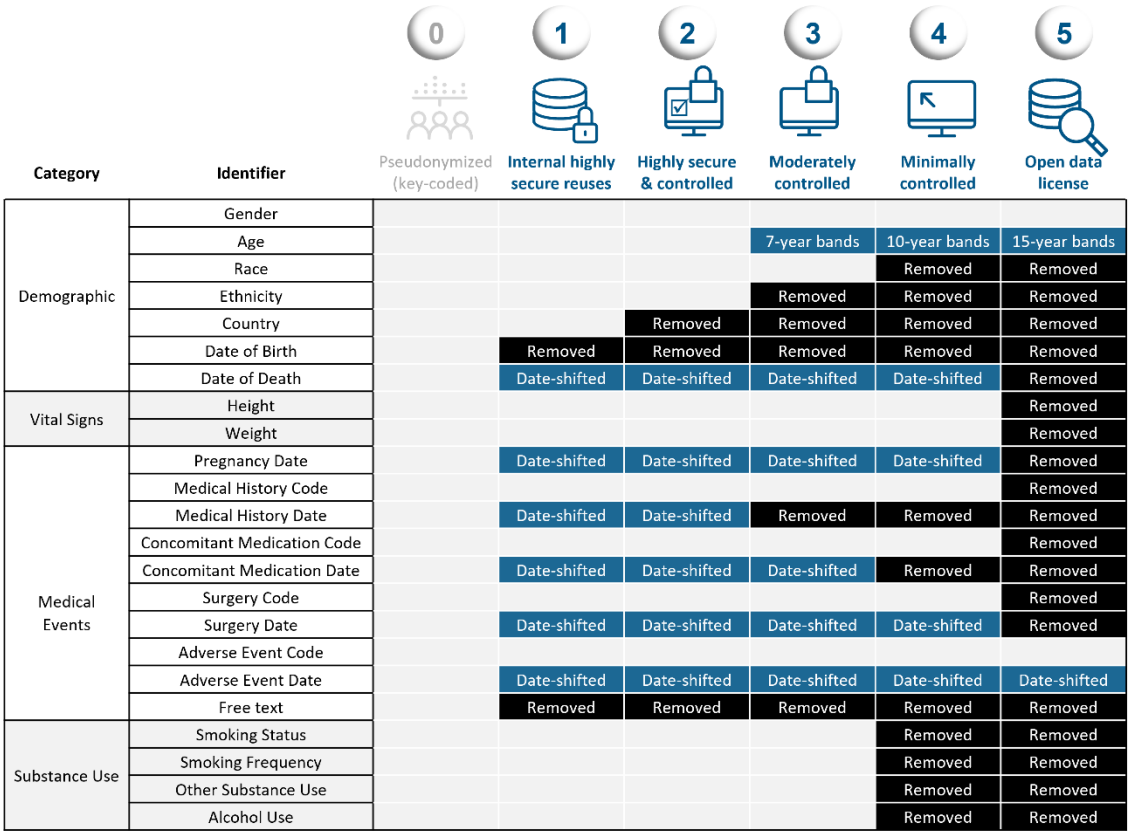


Figure 6 – SAFE Data Standard scale reflecting the degree of data transformation to achieve privacy protection.

Once data are transformed as part of the deidentification and protection process, sponsors should retain reports detailing the approach taken and associated justifications for auditability.

To illustrate the concept of the SAFE Data Standard, we applied the data tolerances from level 1 to 5 to simulate the transformation impacts on indirectly identifying data from a clinical study. (See Table 3 below.) For clarity, directly identifying information or unique identifiers are masked or removed (for instance, such subject IDs are replaced with pseudonyms and site IDs are removed) and non-identifying information, such as a blood glucose reading (which can change frequently), is preserved during the deidentification and protection process.

While the individual variable-level transformations will depend on the study characteristics in practice, including how distinguishable the participants are in the defined population and what preferences for data utility are incorporated (eg, country may be generalized to continent), the general trend of greater transformation and lower utility as you progress from level 1 to 5 is consistent. Table 3 summarizes the results for the simulation performed, providing an example (not a ruleset) of how the SAFE Data Standard can be applied in practice across a range of disclosure contexts.



Category	Identifier	0 Pseudonymized (key-coded)	1 Internal highly secure reuses	2 Highly secure & controlled	3 Moderately controlled	4 Minimally controlled	5 Open data license
Demographic	Gender						
	Age				7-year bands	10-year bands	15-year bands
	Race					Removed	Removed
	Ethnicity				Removed	Removed	Removed
	Country			Removed	Removed	Removed	Removed
	Date of Birth		Removed	Removed	Removed	Removed	Removed
	Date of Death		Date-shifted	Date-shifted	Date-shifted	Date-shifted	Removed
Vital Signs	Height						Removed
	Weight						Removed
Medical Events	Pregnancy Date		Date-shifted	Date-shifted	Date-shifted	Date-shifted	Removed
	Medical History Code						Removed
	Medical History Date		Date-shifted	Date-shifted	Removed	Removed	Removed
	Concomitant Medication Code						Removed
	Concomitant Medication Date		Date-shifted	Date-shifted	Date-shifted	Removed	Removed
	Surgery Code						Removed
	Surgery Date		Date-shifted	Date-shifted	Date-shifted	Date-shifted	Removed
	Adverse Event Code						
	Adverse Event Date		Date-shifted	Date-shifted	Date-shifted	Date-shifted	Date-shifted
	Free text		Removed	Removed	Removed	Removed	Removed
Substance Use	Smoking Status					Removed	Removed
	Smoking Frequency					Removed	Removed
	Other Substance Use					Removed	Removed
	Alcohol Use					Removed	Removed

Figure 7 – Example simulation to illustrate how the SAFE Data Standard can be applied in practice to clinical study data, recognizing that this is an example and not a rule (ie, actual transformations in practice will depend on study characteristics, such as how distinguishable the participants are statistically).

The simulation summarized in Table 3 was based on a randomized, double-blind diabetes study sponsored by Janssen, and Janssen has since made the deidentified data available for secondary research through The YODA Project.[29]

Conclusion

In this response, we have presented the SAFE Data Standard, a framework for sharing deidentified and protected clinical trial data. The SAFE Data Standard draws on a deidentification approach that has been proven effective over years of use and that is sufficiently mature to support the pilot effort described in the RFI. The Standard also provides a method for maximizing the utility of deidentified data by outlining levels of data transformation appropriate to various data sharing contexts. More details on the Standard can be found in the full article in Applied Clinical Trials [2].

We wish to thank you again for this opportunity to provide our views on data collection for emergency clinical trials. We hope that you have found our feedback helpful and insightful towards developing a strategy for conducting large-scale clinical trials in response to future health crises. We look forward to participating in future consultations.

References

1. Seastedt KP, Schwab P, O'Brien Z, Wakida E, Herrera K, Marcelo PG, Agha-Mir-Salim L, Frigola XB, Ndulue EB, Marcelo A, Celi LA. Global healthcare fairness: We should be sharing more, not less, data. PLOS Digital Health. 2022 Oct 6;1(10):e0000102. Available from: https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000102 [accessed Dec 22, 2022].

2. Bamford S, Lyons S, Arbuckle L, Chetelat P. Sharing anonymized and functionally effective (SAFE) data standard for safely sharing rich clinical trial data. Applied Clinical Trials. Available from: https://www.appliedclinicaltrialsonline.com/view/sharing-anonymized-and-functionally-effective-safe-data-standard-for-safely-sharing-rich-clinical-trial-data [accessed Dec 22, 2022].

3. PHUSE Data Transparency Working Group. Data anonymisation and risk assessment automation. PHUSE; 2020 p. 28. Report No.: WP-045. Available from: https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Data+Transparency/Data+Anonymisation+and+Risk+Assessment+Automation.pdf [accessed Dec 22, 2022].

4. Office for Civil Rights, Department of Health and Human Services. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Available from: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [accessed Dec 22, 2022].

5. Polonetsky J, Tene O, Finch K. Shades of gray: seeing the full spectrum of practical data de-identification. Santa Clara Law Review. 2016;56:593–629.

6. Hintze M, El Emam K. Comparing the benefits of pseudonymisation and anonymisation under the GDPR. Journal of Data Protection & Privacy. 2018 Dec 1;2(1):145–58.

7. Barbaro M, Zeller Jr. T. A face is exposed for AOL searcher no. 4417749. New York Times. 2006; Available from: https://nyti.ms/3tKnMUl [accessed Dec 22, 2022].

8. Narayanan A, Shmatikov V. Robust de-anonymization of large sparse datasets. In: Proceedings of the 2008 IEEE Symposium on Security and Privacy. 2008. p. 111–25. doi:10.1109/SP.2008.33

9. Sweeney L. Only you, your doctor, and many others may know. Technology Science. 2015 Sep 29; Available from: https://techscience.org/a/2015092903/ [accessed Dec 22, 2022].

10. Rubinstein I, Hartzog W. Anonymization and risk. Washington Law Review. 2016;91:703–60.

11. PHUSE— the global healthcare data science community. Available from: https://phuse.global/ [accessed Dec 22, 2022].

12. TransCelerate - pharmaceutical research and development. Available from: https://www.transceleratebiopharmainc.com/ [accessed Dec 22, 2022]

13. CRDSA - Clinical Research Data Sharing Alliance. Available from: https://crdsalliance.org/ [accessed Dec 22, 2022].

14. Stadler T, Oprisanu B, Troncoso C. Synthetic data -- anonymisation groundhog day. arXiv:201107018 [cs]. Available from: http://arxiv.org/abs/2011.07018 [accessed Dec 22, 2022].

15. El Emam K, Boardman R. Are clinical trial data shared by the EMA under policy 0070 really public data? Applied Clinical Trials. 2018. Available from: https://www.appliedclinicaltrialsonline.com/view/are-clinical-trial-data-shared-ema-under-policy-0070-really-public-data [accessed Dec 22, 2022].

16. Arbuckle L, El Emam K. Building an anonymization pipeline: creating safe data. Sebastopol, CA: O'Reilly Media; 2020. ISBN:1492053430

17. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS ONE. 2011;6(12). PMID:22164229

18. El Emam K, Arbuckle L. Anonymizing health data: case studies and methods to get you started. Sebastopol, CA: O'Reilly Media; 2013. ISBN:1449363075

19. Jay R, Malcolm W, Parry E, Townsend L, Bapat A. Guide to the General Data Protection Regulation. Sweet & Maxwell; 2017. ISBN:0414061012

20. What is personal data? ICO; 2021. Available from: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/ [accessed Dec 22, 2022].

21. Borgesius FJZ. Singling out people without knowing their names–behavioural targeting, pseudonymous data, and the new Data Protection Regulation. Computer Law & Security Review. 2016;32(2):256–71.

22. PHUSE Data Transparency Working Group. A global view of the clinical transparency landscape—best practices guide. PHUSE; 2020 p. 60. Report No.: WP-35. Available from: https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Data+Transparency/Clinical+Trials+Data+Transparency+Toolkit+Best+Practices+Guide.pdf [accessed Dec 22, 2022].

23. Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. Washington, D.C.; 2015. PMID:25590113

24. Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. N Engl J Med. 2018 Jun 7;378(23):2202–11. PMID:29874542

25. Howe N, Giles EL, Newbury-Birch D, McColl E. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. J Health Serv Res Policy. 2018;23(2):123–33. PMID:29653503

26. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont report. Washington, D.C.: Department of Health, Education, and Welfare.; 1979 Apr. Available from: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html [accessed Dec 22, 2022].

27. European Commission. Ethics and data protection. 2018. Available from: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf [accessed Dec 22, 2022].

28. Kass NE, Natowicz MR, Hull SC, Faden RR, Plantinga L, Gostin LO, et al. The use of medical records in research: what do patients want? J Law Med Ethics. 2003;31(3):429–33. PMID:14626550

29. The YODA Project (NCT01809327). A randomized, double-blind, 5-arm, parallel-group, 26-week, multicenter study to evaluate the efficacy, safety, tolerability of canagliflozin in combination with metformin as initial combination therapy in the treatment of subjects with type 2 diabetes mellitus with inadequate glycemic control with diet and exercise. Available from: https://yoda.yale.edu/nct01809327-randomized-double-blind-5-arm-parallel-group-26-week-multicenter-study-evaluate-efficacy [accessed Dec 22, 2022].

Welcome to Vulcan!

You have received this document because your organization has recently joined Vulcan as a member of the Operations Committee. This document has information that will help you during your onboarding experience and beyond.

Table of Contents

Meet the Vulcan Administrative Team

Darren Weston
dweston6@its.jnj.com
Role within Vulcan: Steering Committee Co-Chair
Ask me about: Why Vulcan matters

Amy Cramer
acramer2@its.jnj.com
Role within Vulcan: Steering Committee Co-Chair
Ask me about: The history of Vulcan, who's involved with Vulcan

Michael van Campen
Michael.vanCampen@HL7Vulcan.org
Role within Vulcan: Vulcan Program Manager
Ask me about: Governance oversight, timelines, outcomes, and quality of Vulcan initiatives and Committees

Stacy Tegan
stacy.tegan@transceleratebiopharmainc.com
Role within Vulcan: Portfolio Manager
Ask me about: Vulcan's connection to TransCelerate

Shani Sampson
shani.sampson@transceleratebiopharmainc.com
Role within Vulcan: Project Manager
Ask me about: Project teams, PMO support, Vulcan events

Hugh Glover
Hugh.Glover@HL7Vulcan.org
Role within Vulcan: Operations Committee Technical Director
Ask me about: Use Case development, project teams, Operations Committee processes

Karen Van Hentenryck
karenvan@HL7.org
Role within Vulcan: HL7 Liaison
Ask me about: HL7 processes, contracting, and invoicing

Confluence

Confluence is the official "single source of truth" platform for HL7 standards development activities and operations. Every work group within HL7 has a space on HL7. You can find the Vulcan Confluence site at this address: https://confluence.hl7.org/display/VA/Vulcan+Accelerator+Home

In order to join the Vulcan Confluence page, you need a free Confluence account. To request an account please fill out the short account request form found here. Your account will be created straightaway and you will be sent a link to set your password. (For any questions or if you encounter any issues, please contact webmaster@hl7.org.)

Once you have created a Confluence account, your access will be updated and you will gain access to Vulcan meeting materials (agendas, slides, presentations, etc.) as well as project development areas.

Vulcan Committees

As of Q1 2022, each committee meets on the following cadence, but this is subject to change:

- **Steering Committee:** Third Tuesday of each month, 2PM – 3PM ET
- **Operations Committee:** First Wednesday of each month, 2PM – 3PM ET
- **Advisory Council:** twice/year
- **Project Teams:**
    1. **Real World Data**: Weekly, Tuesdays at 11AM ET
    2. **Schedule of Activities**: Weekly, Wednesdays at 11AM ET
    3. **Adverse Events**: Weekly, Thursdays at 11AM ET
    4. **Electronic Product Information (ePI)**: Weekly, Thursdays at 6AM ET
    5. **FHIR to OMOP**: Biweekly, Wednesdays at 12PM ET
    6. **Phenotypic Data**: Twice monthly, 1st & 3rd Mondays at 6PM EST

Current Members of Vulcan (as of December 2022)

Organization
Accenture
ACRO
BioVeras
CDISC
Cedars-Sinai Health System
Crohn's & Colitis Foundation
Danish Medicines Agency *
Duke University School of Medicine *
Epic
FDA *
Felleskatalogen
GSK
HL7 International *
IgniteData
Infor
InterSystems
Fujitsu
Johns Hopkins School of Medicine *
Johnson & Johnson
Medidata
Microsoft
MITRE
National Cancer Center Hospital East
NIH National Center for Advancing Translational Sciences *
NIH U.S. National Library of Medicine *
OpenClinica
Oracle
Oregon Health & Science University *
Parexel
Pfizer
PHUSE
Roche
Society of Clinical Data Management *
TransCelerate BioPharma Inc. *
UiO
University of Arkansas for Medical Sciences
University of Colorado Anschutz
UT Health San Antonio Long School of Medicine *
Vanderbilt

*indicates a Convening Member of Vulcan

HL7 Accelerators

The HL7 FHIR Accelerator Program is designed to assist communities and collaborative groups across the global health care spectrum in the creation and adoption of high quality FHIR Implementation Guides or other standard artifacts to move toward the realization of global health data interoperability.

You can find information on other HL7 Accelerators here. Please see below for a brief overview of each existing Accelerator.

Helpful Links

- Getting Started with FHIR
- HL7 Connectathon FAQs
- Participating in HL7
- HL7 Essentials
- How to Use Confluence
- FHIR Dev Days Presentation on Intro to FHIR

Glossary

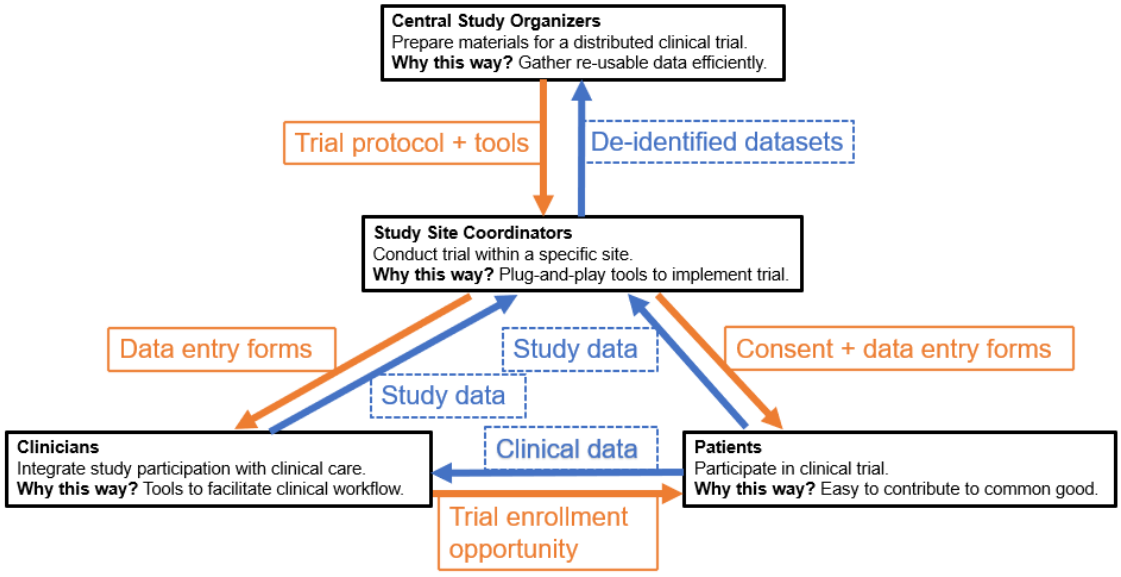
- **Balloting:** The formal process by which proposed standards are reviewed by HL7 members. Participants in the ballot make comments and suggestions about the material presented and may raise objections. Only when all objections have been resolved does the material become an HL7 standard.
- **BR&R:** Biomedical Research and Regulation; a working group within HL7. It is the group most closely aligned with Vulcan and much of Vulcan's formal HL7 work will be channeled through BR&R
- **Connectathon:** An event that is centered on developing the HL7 FHIR Specification including resources, profiles and implementation guides; the purpose of a Connectathon is to prove that the specification is complete and facilitate FHIR implementation guide maturity
- **HL7:** Health Level Seven International
- **IG:** Implementation Guide; a set of rules for using FHIR resources in a particular context
- **FHIR:** Fast Healthcare Interoperability Resources – a means of representing information using standard information building blocks
- **PSS:** Project Scoping Statement – part of the HL7 project management methodology
- **Resources:** Shorthand for **FHIR Resources**. An instance-level representation of some kind of healthcare information building block (e.g., patient, observation); resources are managed by HL7 working groups
- **WGM:** Working Group Meeting; HL7 working group meetings are held three times per year at varying locations or virtually and serve two important purposes: to provide HL7 work groups a chance to meet face to face to work on the standards as well as the opportunity to network with industry leaders from around the world and to provide an invaluable educational resource for the healthcare IT community

OSTP RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot

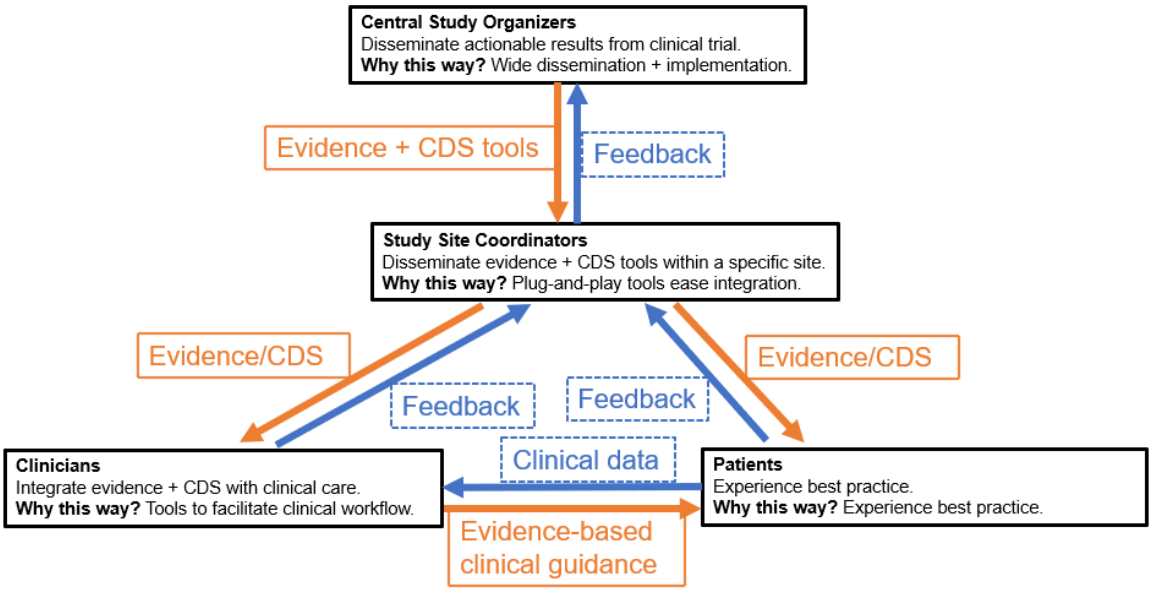
Response from Health Evidence Knowledge Accelerator, a virtual group associated with the Scientific Knowledge Accelerator Foundation. Correspondence to Brian S. Alper, MD, MSPH, President, Scientific Knowledge Accelerator Foundation, balper@computablepublishing.com

The respondent organization is a nonprofit organization with a multidisciplinary constituency including health IT companies, app developers, clinical trial designers, and users of health IT products. The corresponding individual respondent is president of the organization, CEO of a small business that is an IT company providing platform and tooling to support electronic data exchange for scientific knowledge, and project lead for an HL7 project (EBMonFHIR) which is extending FHIR to support data exchange of scientific knowledge including development of FHIR Resources for Citation, Evidence, EvidenceVariable, and ArtifactAssessment.

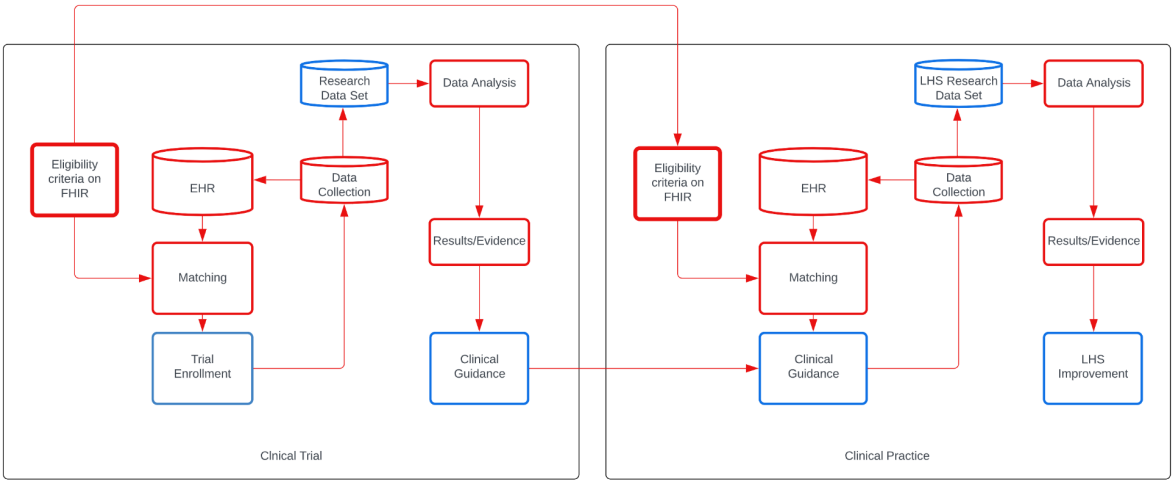
The RFI requests input to support an information sharing model that can be represented as the sharing of knowledge resources between four groups (Central Study Organizers, Study Site Coordinators, Clinicians, and Patients):



The same information sharing model can also be used to facilitate dissemination and implementation of evidence resulting from the clinical trial:



A technical entity diagram showing the similar model for integrating clinical trial conduct with the EHR and integrating clinical decision support with the EHR is:



The model similarities show that the efforts to achieve interoperability for the execution of the clinical trial will also facilitate dissemination and implementation of trial results in subsequent clinical practice.

This RFI response is specific to step 2 of the desired use case:

> "Study sites would enroll participants in the trial (potentially using software mechanisms that can alert sites to potential subjects for a specific protocol in a manner that increases the diversity of trial populations). Sites would obtain appropriate e-consents and authorizations from participants."

Specifically, this RFI response describes how trial eligibility criteria structured in FHIR EvidenceVariable Resources could facilitate this step for a rapid interoperability pilot.

Lacking a simple, re-usable form for the expression of eligibility criteria is a common interoperability challenge for:

- study site selection (based on determination of an adequate number of patients eligible to participate),
- clinical trial recruitment (based on matching eligible patients), and
- the use of trial results for decision support (based on matching eligible patients).

Such eligibility criteria may also be called cohort definitions or clinical phenotypes.

Existing efforts to model structured eligibility criteria include formal expression languages, such as Clinical Quality Language (CQL), and non-FHIR specialized data structures, such as Phenotype KnowledgeBase (PheKB).

The ease of implementation for these efforts has not advanced sufficiently to support an interoperability pilot in 6-12 months. It would take longer to develop and scale tooling and training materials to involve these technologies on such a rapid timeline.

HL7 members, seeking to define eligibility criteria for clinical trials in a FHIR Resource, collaborated with the EBMonFHIR project and we adapted the EvidenceVariable Resource for this use case.

The EvidenceVariable Resource can be used to provide each eligibility criterion in structured form, and the form can be applied with any of the following datatypes:

1) A string, used when structured form is unnecessary or inappropriate
2) A reference to another FHIR Resource where the criterion is defined
3) A canonical URL for a direct link to where the criterion is defined
4) A codeable concept in which the criterion is defined in a structured terminology
5) An expression in which CQL or another expression language can be used
6) A type-and-value combination in which the type of criterion can be expressed with a codeable concept and its value (the values within which the criterion would be met) can be defined with a variety of datatypes (codeable concept, boolean, quantity, range, or reference to another FHIR Resource).
7) A combination of criteria, in which a code can be used to define the method of combination (such as all-of, any-of, at-least, or at-most)

In addition, timing elements can be added to define when any criterion is met.

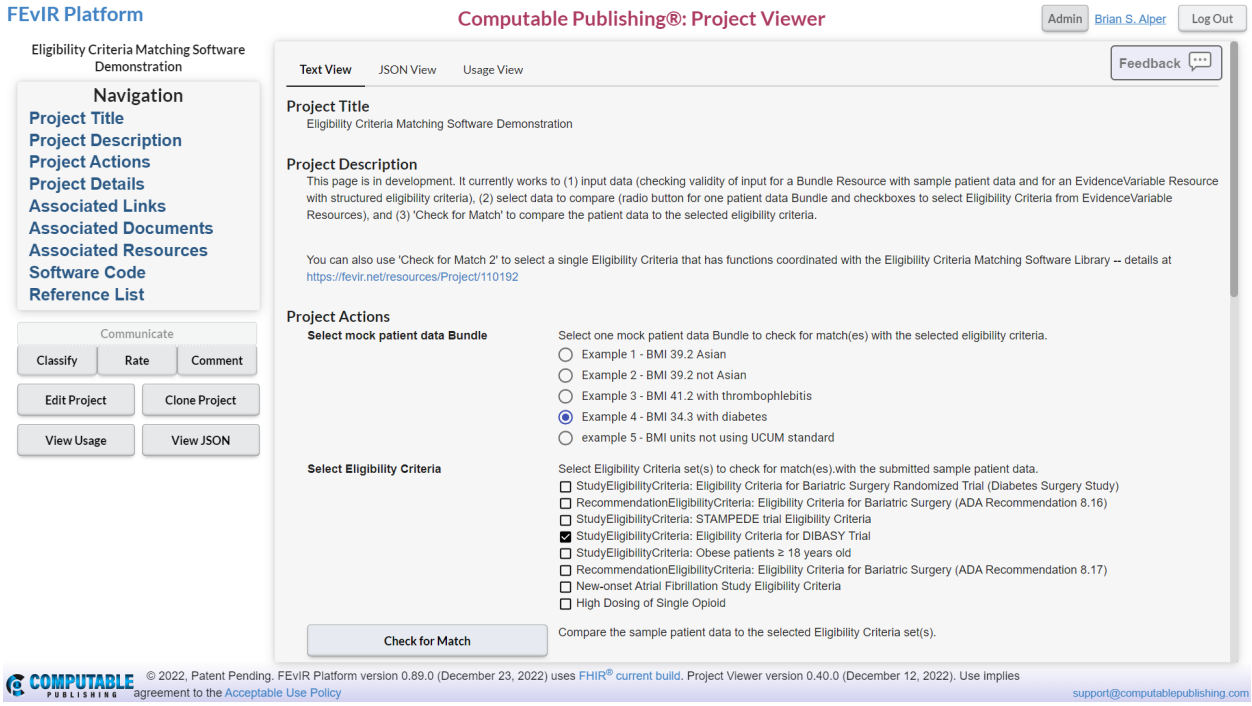
Regardless of the specific datatype used to express eligibility criteria as structured data in a FHIR EvidenceVariable Resource, the ability to use a FHIR-based Resource enables:

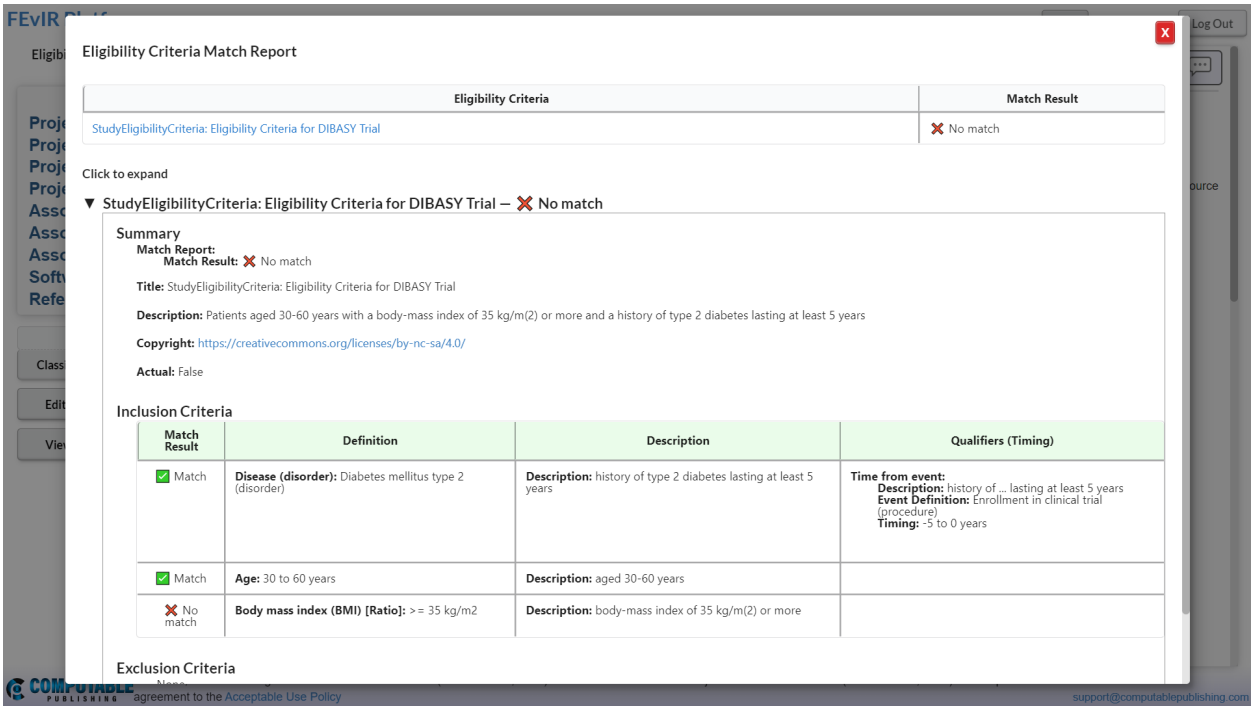
1) Developers already using FHIR can use the same systems for data exchange for eligibility criteria.
2) The tools in development to support human-friendly data entry to express eligibility criteria and automatically convert the data to FHIR EvidenceVariable Resource structure can be completed in time for scaled pilot use in 6 months.
3) Tools to convert FHIR EvidenceVariable expressions of eligibility criteria to CQL expressions can be developed more efficiently than either developing tools to convert natural language data to CQL or training new people to use CQL authoring tools directly.

Three project pages on the Fast Evidence Interoperability Resources (FEvIR) Platform can be viewed to demonstrate these concepts as they are developing.

First, the 'Eligibility Criteria specification with EvidenceVariable' project at https://fevir.net/32444 can be used to view multiple examples of structured eligibility criteria in FHIR EvidenceVariable Resources. When ready we will add links to an Implementation Guide developed by an HL7 project (FHIR Representation of Eligibility Criteria for Clinical Trials at https://jira.hl7.org/browse/PSS-2127) that has been approved and will develop an Implementation Guide in 2023. This effort could be coordinated with an Interoperability Pilot.

Second, the 'Eligibility Criteria Matching Software Demonstration' project at https://fevir.net/51402 can support demonstration of a simple matching algorithm. On this page you can enter a mock patient data bundle (a FHIR Bundle Resource with 1 Patient Resource and any number of Observation and Condition Resources) or select an example mock patient data bundle. You can then enter eligibility criteria as an EvidenceVariable Resource or select an example. Screenshots on the next page will show the result of selecting "Example 4 - BMI 34.3 with diabetes" and "StudyEligibilityCriteria: Eligibility Criteria for DIBASY Trial" and then clicking the "Check for Match" button.


FEvIR Platform
Computable Publishing®: Project Viewer
Admin
Brian S. Alper
Log Out
Eligibility Criteria Matching Software Demonstration
Navigation
Project Title
Project Description
Project Actions
Project Details
Associated Links
Associated Documents
Associated Resources
Software Code
Reference List
Communicate
Classify
Rate
Comment
Edit Project
Clone Project
View Usage
View JSON
Text View
JSON View
Usage View
Feedback
Project Title
Eligibility Criteria Matching Software Demonstration
Project Description
This page is in development. It currently works to (1) input data (checking validity of input for a Bundle Resource with sample patient data and for an EvidenceVariable Resource with structured eligibility criteria), (2) select data to compare (radio button for one patient data Bundle and checkboxes to select Eligibility Criteria from EvidenceVariable Resources), and (3) 'Check for Match' to compare the patient data to the selected eligibility criteria.
You can also use 'Check for Match 2' to select a single Eligibility Criteria that has functions coordinated with the Eligibility Criteria Matching Software Library -- details at https://fevir.net/resources/Project/110192
Project Actions
Select mock patient data Bundle
Select one mock patient data Bundle to check for match(es) with the selected eligibility criteria.
Example 1 - BMI 39.2 Asian
Example 2 - BMI 39.2 not Asian
Example 3 - BMI 41.2 with thrombophlebitis
Example 4 - BMI 34.3 with diabetes
example 5 - BMI units not using UCUM standard
Select Eligibility Criteria
Select Eligibility Criteria set(s) to check for match(es).with the submitted sample patient data.
StudyEligibilityCriteria: Eligibility Criteria for Bariatric Surgery Randomized Trial (Diabetes Surgery Study)
RecommendationEligibilityCriteria: Eligibility Criteria for Bariatric Surgery (ADA Recommendation 8.16)
StudyEligibilityCriteria: STAMPEDE trial Eligibility Criteria
StudyEligibilityCriteria: Eligibility Criteria for DIBASY Trial
StudyEligibilityCriteria: Obese patients ≥ 18 years old
RecommendationEligibilityCriteria: Eligibility Criteria for Bariatric Surgery (ADA Recommendation 8.17)
New-onset Atrial Fibrillation Study Eligibility Criteria
High Dosing of Single Opioid
Check for Match
Compare the sample patient data to the selected Eligibility Criteria set(s).
© 2022, Patent Pending. FEvIR Platform version 0.89.0 (December 23, 2022) uses FHIR® current build. Project Viewer version 0.40.0 (December 12, 2022). Use implies agreement to the Acceptable Use Policy
support@computablepublishing.com


Eligibility Criteria Match Report
Eligibility Criteria
Match Result
StudyEligibilityCriteria: Eligibility Criteria for DIBASY Trial
No match
Click to expand
▼ StudyEligibilityCriteria: Eligibility Criteria for DIBASY Trial — No match
Summary
Match Report:
Match Result: No match
Title: StudyEligibilityCriteria: Eligibility Criteria for DIBASY Trial
Description: Patients aged 30-60 years with a body-mass index of 35 kg/m(2) or more and a history of type 2 diabetes lasting at least 5 years
Copyright: https://creativecommons.org/licenses/by-nc-sa/4.0/
Actual: False
Inclusion Criteria
Match Result
Definition
Description
Qualifiers (Timing)
Match
Disease (disorder): Diabetes mellitus type 2 (disorder)
Description: history of type 2 diabetes lasting at least 5 years
Time from event:
Description: history of ... lasting at least 5 years
Event Definition: Enrollment in clinical trial (procedure)
Timing: -5 to 0 years
Match
Age: 30 to 60 years
Description: aged 30-60 years
No match
Body mass index (BMI) [Ratio]: >= 35 kg/m2
Description: body-mass index of 35 kg/m(2) or more
Exclusion Criteria

Third, the ‘Eligibility Criteria Matching Software Library’ project at https://fevir.net/110192 provides shareable code (JavaScript functions) to support re-usable criteria expression and criteria matching expressions. This project was initially created to demonstrate simple functions without integration with CQL or other scaled systems, but the project can be adapted to share code and expressions using languages other than JavaScript.

We have also used FHIR EvidenceVariable (for the eligibility criteria for a research study), ResearchStudy, ResearchSubject, and Consent Resources to facilitate online enrollment and consent for a non-clinical study. In this context, we provided a human-friendly display of the eligibility criteria for potential participants to confirm their eligibility, then provided the consent document as an attachment for viewing, and upon electronic confirmation of consent, created Consent and ResearchSubject Resources to use for documentation and subsequent research conduct. The original project was ‘Risk of Bias Assessment Tool (RoBAT) Usability Research (RoBATUR)’ at https://fevir.net/29571 but that project is closed so you cannot view the enrollment process directly. We later created a ‘Risk of Bias Assessment Tool (RoBAT) Usability Research (RoBATUR) 2’ project at https://fevir.net/52377 which is inactive but you can view it to experience the enrollment process.

These projects are advanced through the Health Evidence Knowledge Accelerator, a virtual group with 14 active weekly working group meetings. The meeting schedule can be found at https://fevir.net/29272 and these activities can be adapted to coordinate with an Interoperability Pilot.

Request for Information (RFI) Response Form

In response to the Request for Information

By: Office of Science and Technology Policy (OSTP)

For: Data Collection for Emergency Clinical Trials and Interoperability Pilot

Reference: 2022-23489

Date of this Response: December 27, 2022

SECTION 1: About the Respondent

This is a Response by Acoer, Inc (the Respondent) to provide information.

Item	Detail
Organization name:	Acoer, Inc
Stakeholder type:	health information technology (health IT) company & app developers
Physical address:	Marietta, GA
Business website:	https://www.acoer.com/
Business email:	Tech@acoer.com

SECTION 2: Responses

2. HL7 FHIR APIs. We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-emergency phase, and in what areas further advances might be needed. Specific topics in this connection include:
a. Use of an API that supports FHIR Bulk Data Access to support clinical research; whether bulk data exports from EHR systems can be used to support certain clinical trial protocols.

Acoer leverages FHIR APIs within our HealthReady application in order to confidently and securely sync patient data which can then be pseudonymously available for clinical trial researchers.

Patient data access is managed via integration with technology partner 1up Health (or other EHR/FHIR syncing solutions), which provides an open API gateway—using Fast Healthcare Interoperability Resources (FHIR)—to provide secure access to patient data stored on native Electronic Medical Record (EMR) systems. The patient has full control over this process, as they search for their provider and enter their confidential login information for that chosen provider in order for their records to successfully sync. Acoer also uses a FHIR Subscription Listener which allows the patients' healthcare data to automatically sync, as long as consent permits, so their records are always up to date within the HealthReady application. This data is then pseudonymously available for clinical trial researchers to find patients they would like to participate. The HealthReady application provides tools for each patient (user) to access and monitor their patient record transactions and utilize their data for potential clinical research.

As patient healthcare records are synced automatically through the FHIR APIs once consented by the patient, clinical researchers are able to log into the researcher portal and feel confident that they are filtering up to date information, especially in the event of an emergency, and can instantly request participation from the selected patients that fit their criteria.

6. Consent, deidentification, return of results. The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution.
a. In light of this, we seek comment on how the tools described above can be used to obtain, collect and/or manage any required informed consents and/or authorizations from patients or individuals in accordance with applicable regulations.

The HealthReady application uses Hedera Hashgraph Distributed Ledger Technology (DLT) to capture consent in a deidentifying and decentralized way. Consents are created and signed using Non-Fungible Tokens (NFTs) to represent and manage the patients' consent and its rights and protections. This is important for the patient so that the individual owns and controls the consent and is also able to monitor its use. NFTs through Hedera Hashgraph provide:

- Accountability through cryptographic and real-time proof of action
- Security through decentralized processing and storage
- Trust through an automated, continuous and transparent auditing of transactions

b. We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information. It would be helpful to receive comments on which deidentification and protection approaches are sufficiently mature to support a pilot effort in the near term.

As the patient consents through NFTs on Hedera Hashgraph and syncs their patient data, the researchers will then be able to filter and search through anonymized data and metadata that does not risk patient deidentification providing "aggregated results" such as # of patients, # of medical resources, classification by age group, etc. that fits within their clinical trial criteria.

c. Ideally, patient authorization would allow clinical trial data to be used for additional research beyond the original study. We would appreciate input on how the content collected for consent and authorization as well as the interfaces with deidentification technologies should be designed to enable flexible and responsible reuse of clinical trial data.

As the patient consents through NFTs on Hedera Hashgraph, all pseudonymized data will be available as long as the patient consent is active. This allows researchers to have continuous access to available data and the ability to request participation to all patients within the HealthReady platform, one or many times, as they fit clinical trial criteria.

e. We seek comment on any regulatory or ethical guidelines that are relevant to patients' consents and authorizations under the use case described in this RFI, and on ways in which technical solutions might help ensure adherence to applicable regulatory or ethical guidelines.

In order to ensure adherence to applicable regulator or ethical guidelines, a technical solution called Data Stamping has been integrated within the HealthReady platform. Data Stamping is an API created by Acoer which uses Hedera Hashgraph's Hedera Consensus Service (HCS) that

has the ability to stamp any digital asset or action and record it on the ledger. The metadata stamped contains a standardized group of data such as, geolocation, date and time, expiration, action taken, etc. based on healthcare and clinical trial guidelines. These stamps provide the validity of proof including duration, signature and ledger transaction as well as an audit trail of actions, updates, or changes to the digital asset ensuring there are no bad actors acting within the platform.

7. User interface and experience. With all of the above technologies, we seek input on:
a. The best way to optimize the experience of health care providers, administrators, and other users, so as to maximize the utility and uptake of the product.

As our product focuses on direct patient to researcher clinical trial participation through the FHIR syncing of the patients' EHR, our user interface and experience within HealthReady reflects that focus. HealthReady has three portals within the platform - the Patient Portal, the Researcher Portal, and the Administrator Portal.

The Patient Portal provides an incentivization mechanism to enable compensation for patients in contributing their personal health data to clinical research. This incentivization is done through tokenization through HealthReady minting HR Tokens on Hedera Hashgraph using Hedera Token Service (HTS) and Hedera's Smart Contract 2.0 (SC2). These minted tokens will be transferred to patients (using HTS/SC2) for the following actions:

- Syncing of their medical recorders
- Selected by researcher for CT
- Acceptance of participation of CT

Patients (users) can then exchange their tokens for reward redemption (1HR token = 1USD) using HTS/SC2. Patients can also feel confident that all clinical trial data within the application is up to date and can feel confident doing their own research for clinical trials to participate in as clinical trials are synced daily to the application.

The Researcher Portal provides researchers a clean, visual and easy way to search for qualified participants that fit their clinical trial criteria. The HealthReady platform provides filters on a number of relevant pseudonymized criteria (such as condition, gender, age group, etc.) to deliver the most relevant user experience. The platform also manages the request and quote process between researcher and patient once the researcher selects their qualified patients. The platform will then display contact information provided by AACT if patient signifies interest in joining the clinical trial.

The Administrator Portal provides HealthReady the ability to monitor users across all platforms as well as the number of quotes from researchers to patients, the number of payments to patients due to participation within clinical trials, and general maintenance items.

c. For clinicians and health IT users: what existing tools, apps, or processes you have found most usable and why.

1UpHealth is a solution we have found usable as they provide the ability to sync patient's healthcare records within the HealthReady application through FHIR APIs and eliminate the need for users to have to manually enter all of their health records data. Also, once the user has synced and if their consent is still active, all new and updated information will automatically sync to the user's account.

10. Emerging technologies. We welcome comments on any future technological developments we should anticipate. Relevant technical developments include but are not limited to differential privacy; federated machine learning; other technologies referenced in the recent OSTP RFI related to privacy-enhancing technologies (PET) (seeFederal Register: Request for Information on Advancing Privacy-Enhancing Technologies); and technologies outside of the PET space. Specific topics in this area include:
a. How future technologies might affect the use case and underlying assumptions laid out in this RFI.

Distributed Ledger Technology (DLT) needs to be highly considered as a critical component in the future of healthcare innovations. DLT provides secure data management through decentralized nodes which allow for clear tracking and auditability of assets. Acoer's data stamping APIs provide the ability to stamp any type of asset on the Hedera Hashgraph ledger to track any metadata such as date, name, size, contents of asset (through hashing), etc. It is important to note that no critical or personal information is available on the public ledger; however, it provides computational trust in the asset itself, whether consent, health records, billing records, etc.

b. How future technologies might change the nature of the software architecture, data architecture, or potential data collection solutions for clinical trials.

As discussed throughout this RFI, the HealthReady platform uses hedera Hashgraph DLT technology to promote a secure, trustworthy yet transparent location for patients and researchers to come together and meet the needs both groups have – the need to find and participate in clinical trials meets the need to find participants which meet their clinical trial criteria.

11. Pilot or demonstration project. We seek comment on how the U.S. Government can best work with external stakeholders and developers to develop a pilot or demonstration project that will operationalize clinical trial data capture and serve as a basis and model for data collection in the event of an emergency. This pilot or demonstration project could also potentially support clinical research in the pre-emergency phase. Specific topics include:
a. Whether data can be managed through a central repository or small set of central data repositories; options for cloud-based data storage.

As the HealthReady platform, its technology, architecture, and methods have been discussed throughout this RFI, we also would like to propose a pilot project with the U.S. government

where we can make use of HealthReady and expand as needed, or produce an additional pilot project using the DLT technology put forth in order to meet the U.S. governments specific specifications on this subject matter. Our proposal would include having the data stored and managed through cloud storage where no patient data is stored, only required information for discovery, analytics and exploration. We would also propose the use of decentralized DLT approaches for Data Stamping for asset and consent management to provide the ability to generate validity of the proof for authenticity for complete trust and transparency.

b. Technical options that might hold promise in the short term to enable researchers from diverse locations to analyze the data collected from multiple clinical trial sites. We also seek comment on any additional options that should be considered in the long term.

We currently provide clinical trials synced daily from clinicaltrials.gov making them available to patients (users) and researchers, however, we have the ability to sync from other clinical trials sites to expand the data collection for patients and researchers alike.

c. Whether any parts of the pilot would be appropriately supported as
i. A demonstration project with commercial partnership.

Acoer would be happy to discuss a demonstration project with a commercial partnership where Acoer's technology can be used within another entity's project to help support Data Collection for Emergency Clinical Trials and Interoperability. The implementation of a blockchain solution would be highly beneficial for healthcare companies to provide a layer of trust with real-time data and in depth audit trails to ensure the clinical trial solutions remain ethical, secure and protected.

ii. A public-private partnership.

A public-private partnership would be an ideal situation for this pilot project. Acoer has the technical expertise in the blockchain-healthcare space with proven solutions within this realm. Partnering Acoer with a knowledgeable, public entity such as the Office of Science and Technology Policy (OSTP) as the subject matter experts (SME), to provide specifications for an ideal platform to tailor our technology solutions for solving the need for Data Collection for Emergency Clinical Trials and Interoperability would allow for a rich, intuitive, real-time and computationally trusted environment for its users and researchers.

iii. An agency-funded program.

An agency-funded program could also provide the needed funding and subject matter experts in the area of clinical trial research where Acoer can continue in its technological growth with HealthReady architecture and technology to improve the patient and researcher relationship within the Clinical Trial space to provide a demonstrable pilot project to help provide a solution for this Data Collection for Emergency Clinical Trials and Interoperability Pilot.

12. Specific commercial capabilities. Commenters who are developing a technology or product that might be relevant to any of the topics set forth above are welcome to include a description of that product. Comments about a specific technology or product should be limited to three pages or less.

Patients and advocacy groups are critical to clinical research projects yet there is little transparency, engagement or incentivization to get their insights and/or participate in the clinical study design and research process. This is particularly damaging to rare disease research. Our vision is transparency for transaction flow and incentivization to engage all parties equally including patients, advance groups, clinical researchers, CROs, sponsors, investigators and other relevant stakeholders.

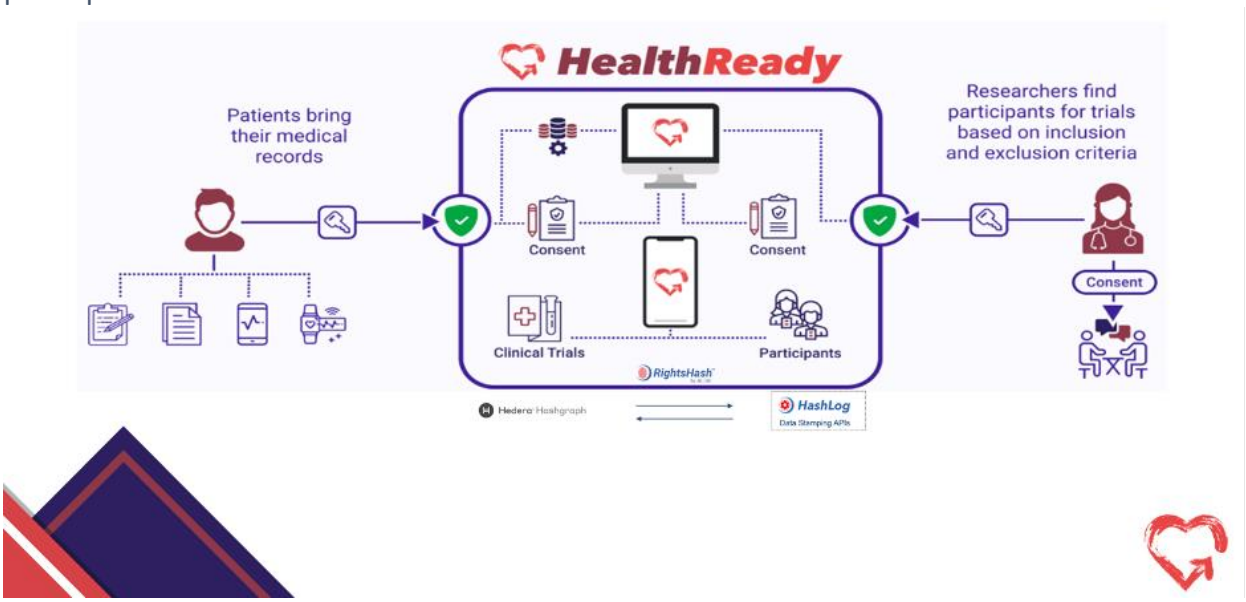
Our solution, HealthReady, provides a platform that stores HIPAA-compliant demographic data and allows researchers to directly purchase anonymized data from patients. When the purchase of a full dataset is initiated, our rules engine matches the demographic data to its corresponding medical data. This process links the demographic dataset and its privacy preserving hashed reference, in addition to initiating payment to the patient's wallet (for example using HBARs). This type of exchange has the power to both reduce the cost of research and return ownership of health data to patients.

Key benefits our solution will provide:

- Empower patients to control and access their data (EMR/EHR, Health Apps, Advocate groups, etc.) and decide who and what data is shared
- Incentivize and reward patient data sharing, while ensuring trustworthiness and control over the data stream
- Address inefficiency of drug development process for patient populations with high unmet patient needs by bringing new drugs early to market and using patient health data consented for:
    - Real World Evidence (RWE) analysis
    - Safety/Efficacy assessment
    - Recruitment into clinical trials
    - Research of the disease and other unmet needs particularly rare diseases
    - Study design
- Provide a longitudinal view of the patient's health journey by enabling permissioned access to patient information for healthcare professionals to better assess patient's health, identify the best course of treatment, and better understand the patient's disease
- Provide a mechanism for real-time, secure, transparent, and auditable collaboration in clinical research amongst a group of peers
- Data transparency for patients, healthcare practitioners, and regulatory authorities to access information in real-time, enabling faster submission for drug applications

Network Lifecycle Diagram

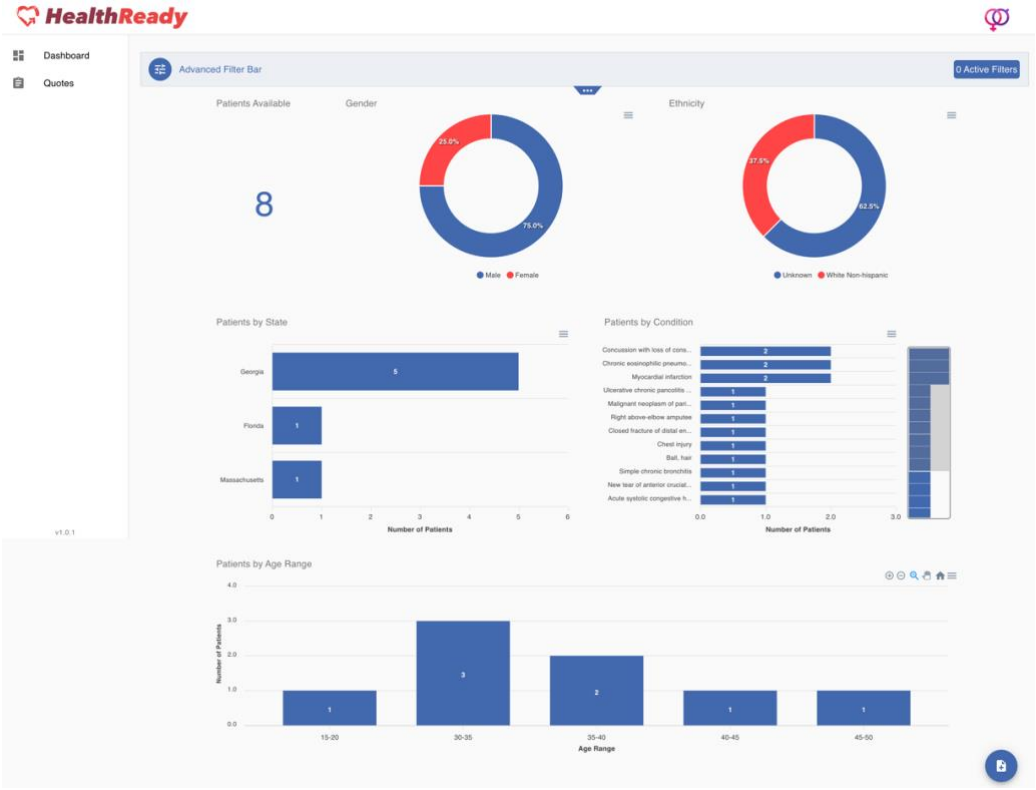
The diagram below shows at the high-level how our platform provides effortless matching of participants to clinical trials.



Patient App Demo

https://www.youtube.com/watch?v=XGxXbKeGqaQ

Screenshot of demonstration Researcher Portal



ENVEIL

ENCRYPTED VEIL

Data Collection for Emergency Clinical Trials and Interoperability Pilot

RFI Response

Prepared for: Office of Science and Technology Policy

Point of Contact:
Lisa L. Bader
Chief Marketing Officer
lisa@enveil.com / 970.988.6475

Introduction

As a Privacy Enhancing Technology solution provider that has operated in the commercial and government space for six years, Enveil is uniquely positioned to provide insight into how OSTP can best operationalize clinical trial data capture in a secure and private manner. For context, Enveil's NIAP/CSfC-certified ZeroReveal® COTS solutions enable Trusted Compute in Untrusted Locations® by allowing data to be securely processed at scale while remaining in the untrusted domain. Defining the transformative category of Privacy Enhancing Technologies (PETs), Enveil ensures the content of the search, analytic, or machine learning model – and its corresponding results – remain encrypted during processing. Whether performing searches or analytics on data within an organization's walls or seeking information from a third-party data provider, Enveil ZeroReveal® ensures nothing is exposed during the entire processing lifecycle.

In this response to OTSP's Request for Information, our goal is to offer insights on technical strategies to support optimized data collection for U.S. clinical trials carried out across a range of institutions and sites, both in emergency settings and in the pre-emergency phase.

Responses to Select Topic Areas

10. Emerging technologies. We welcome comments on any future technological developments we should anticipate. Relevant technical developments include but are not limited to differential privacy; federated machine learning; other technologies referenced in the recent OSTP RFI related to privacy-enhancing technologies (PET) ; and technologies outside of the PET space.

Solutions representing the Privacy Enhancing Technologies (PETs) category, which includes technologies such as homomorphic encryption and secure multiparty computation, address a previously unmet need by allowing entities to securely search and derive insights from untrusted or third-party data assets in a decentralized manner without revealing the contents of the search itself or compromising the security/ownership of the underlying data. By keeping sensitive search terms, analytics, and machine learning models protected throughout the entire processing lifecycle, PETs allows users to securely derive insight from multiple decentralized data sources, even when using highly sensitive or regulated data. Enveil has worked diligently over the years to educate the market on the ways in which these commercially practical, market-ready solutions can enable a range of critical capabilities, such as secure and decentralized data usage for scenarios such as those outlined in this Emergency Clinical Trials and Interoperability Pilot.

At its core, PETs are a family of technologies that enable, enhance, and preserve the privacy of data throughout its lifecycle; they are increasingly gaining attention for the critical role they play in enabling

cross-border, cross-sector, and cross-silo collaboration. The technologies within this category uniquely provide essential privacy assurances by allowing disparate entities to securely derive insights from decentralized data sources without pooling, standardizing, or compromising the security or ownership of the underlying data. Within the health sector, PETs power interoperability while maintaining privacy and accuracy to help researchers gain deeper and more meaningful insights. Healthcare professionals, researchers, suppliers, and other stakeholders can securely share and collaborate with critical data – even sensitive clinical data — without compromising sensitive search terms or the privacy of the underlying data.

PETs-based solutions, such as Enveil ZeroReveal®, have the potential to change the nature of the approach, architecture, and data collection strategy for large-scale clinical trials involving multiple institutions and sites, such as the example outlined in this RFI. For example, when PETs are used specifically to protect the interaction with the data (such as the query, analytic, or machine learning model), collaboration can take place in a decentralized manner that allows all contributors to maintain positive control and ownership of their data assets and without needing to send the data to a central repository for researchers to analyze. This uniquely allows third-party entities or users at different data collection sites to securely and privately work together in ways that were never before possible. By protecting data while it's being used or processed and without requiring data to be moved from its point of origin, PETs can enable OSTP to share large-scale clinical trial data quickly and efficiently without moving that data from the point of collection.

As a pioneering PETs provider, Enveil's ZeroReveal ground-breaking solutions provide secure, decentralized collaboration capabilities for customers in both the government and commercial space. The following outline offers insight into our two main product lines, ZeroReveal Search and ZeroReveal Machine Learning (additional detail is provided in our response to Topic Area #12, below).

Enveil ZeroReveal® Search provides the ability to take searches containing sensitive content, encrypt them, and then run them without ever decrypting them in the untrusted data environment and thus without ever exposing any of the sensitive content outside of the requesting party's trusted walls. ZeroReveal Search also enables encrypted watchlisting for the purpose of encrypted tipping and alerting. As a situationally relevant example, OSTP could 'park' homomorphically encrypted watchlists with indicators of importance (specific results, screening characteristics, etc.) in untrusted locations, such as the various collection sites. As data streams past the watchlist, encrypted tips and alerts are generated greatly reducing analyst workload and ensuring that OSTP is aware of the critical result as soon as it is collected. The acceleration of information flow could also reduce the transfer of superfluous data, reducing storage costs and time to insight. These search and watchlisting capabilities can be performed over any type of data, including textual data and non-textual data such as imagery and biometrics.

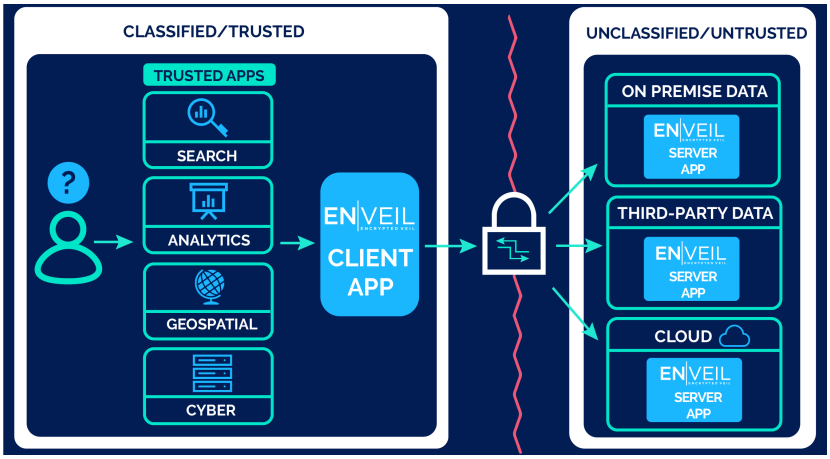
Engineering trust through the power of Privacy Enhancing Technologies (PETs), **Enveil's ZeroReveal® Machine Learning** solutions extend the boundary of trusted compute by enabling encrypted federated learning and secure data usage across disparate, decentralized datasets for machine learning applications.

- Encrypted Evaluation: Powered by homomorphic encryption, ZeroReveal ML Encrypted Evaluation ensures ML models and their associated results remain encrypted throughout the entire evaluation lifecycle. Organizations can securely and privately derive insights from data sources across jurisdictional, third-party, and organizational boundaries, even when using highly sensitive or proprietary models, including those trained on sensitive data.

- Encrypted Training: Utilizing advances in PETs, namely secure multiparty computation (SMPC), ZeroReveal ML Encrypted Training (ZMET) allows organizations to train models in an encrypted capacity. This encrypted training process enables secure federated learning, protecting the model development process, the data used for training, as well as the interests and intent of the parties involved. Users can confidently leverage sensitive data and/or ML models during training, resulting in enhanced models that can more accurately derive insights and deliver value.

ZeroReveal Architecture and Deployment

Enveil ZeroReveal is a lightweight, API-based proxy-layer software system ready for deployment and immediate mission impact. By decoupling from the storage technology layer, Enveil sits above the data, requiring no changes to the underlying environment and easily integrating with existing authentication and audit systems. As a two-party, proxy-layer software system Enveil ZeroReveal consists of (1) the ZeroReveal® Client application and (2) the ZeroReveal® Server application.



Typical Federal Deployment Architecture Diagram

The **ZeroReveal Client application** lives in the trusted environment and integrates via API with existing user interfaces, workflows, authentication, and audit frameworks. It can be deployed anywhere the customer designates, from user-environments to the enterprise service level. The Client application encrypts the operations (searches, analytics, and ML models) before they leave the trusted environment and decrypts the encrypted corresponding results upon receipt.

The **ZeroReveal Server application** is deployed on or near the data in the existing storage architecture and integrates with existing authentication and audit frameworks. The Server application receives the encrypted searches, analytics, and ML models and processes them over the data assets to which it has been granted access without ever decrypting them in the data owner's environment. The Server application sits above the data layer, meaning Enveil can seamlessly integrate with both common and non-standard data storage and database technologies.

Both the ZeroReveal Client and Server applications are highly configurable and can be leveraged for one-to-one, one-to-many, many-to-one, and many-to-many deployments, enabling analysts to securely use data that they have access to as it exists at any other location, regardless of security fabric. ZeroReveal is proven to work robustly and at speed over terabytes of data on cloud platforms and in third-party data locations.

11. Pilot or demonstration project. We seek comment on how the U.S. Government can best work with external stakeholders and developers to develop a pilot or demonstration project that will operationalize clinical trial data capture and serve as a basis and model for data collection in the event of an emergency. This pilot or demonstration project could also potentially support clinical research in the pre-emergency phase. By keeping sensitive search terms, analytics, and machine learning models protected throughout the entire processing lifecycle, PETs allow users to securely derive insight from multiple data sources, even when using highly sensitive or regulated data. Through its ZeroReveal® solutions, Enveil uniquely provides a decentralized approach to secure data collaboration, allowing participating entities to retain granular control and ownership of their sensitive assets. Since requirements to move or pool sensitive assets frequently prove to be the breaking point in collaboration efforts, Enveil removes that risk by allowing organizations to securely and privately search and share over cross-silo and third-party data *as it is* and *where it is* today. These capabilities can enable researchers from diverse locations to securely analyze the data collected from multiple clinical trial sites without needing to move the data from the point of collection. In a crisis situation, eliminating the need to move the data would enable faster access to the data and would also negate the stagnant data issue, as there would be no need to constantly update a centralized repository.

To paint a picture of what this type of PETs-powered decentralized collaboration might look like, consider an example in the financial services industry, specifically, a global bank's customer onboarding process. If a UK bank wants to check information relating to a possible new customer during onboarding, there is no efficient privacy-preserving, compliant way for them to ask other branches across the globe if they know anything about the customer being considered. Using PETs, the bank could encrypt a search in the UK jurisdiction where the customer is trying to open an account, and then send that encrypted search to run against customer databases in other jurisdictions. The encrypted search is performed without ever exposing the customer information, and the encrypted result is returned to the onboarding jurisdiction. Only then is the information decrypted and viewable to the analyst. Because PETs allow that search to be processed while encrypted, customer privacy is maintained and regulations are respected, giving the bank a broader and clearer operating picture without introducing additional risk.

Beyond financial services, there are many ways PETs-powered, decentralized collaboration could enable business functions across verticals: supporting sensitive data sharing in the healthcare industry for efforts like large-scale clinical trials, facilitating a data marketplace where partners could contribute to and access consumer transaction records in a secure and private way, allowing industry groups to benchmark trends without exposing individual contributors, and enabling manufacturers to monitor equipment performance statistics without compromising customer usage patterns. And, thanks to recent breakthroughs in the utilization of PETs, all of this can happen in a matter of seconds.

As OSTP seeks viable technical strategies to distribute clinical trial protocols and capture clinical trial data in order to operationalize protocol distribution and data capture, we recommend a pilot or demonstration project utilizing a decentralized approach to the challenge outlined in the RFI. As described in this response, PETs-based solutions such as Enveil ZeroReveal provide unique decentralized collaboration capabilities that can enable OSTP to share large-scale clinical trial data quickly and efficiently without moving that data from the point of collection. This approach has the potential to transform the ability of U.S. clinical trials enterprise to carry out coordinated, large-scale research protocols in an emergency setting thereby enlarging and strengthening the overall clinical trials infrastructure.

12. Specific commercial capabilities. Commenters who are developing a technology or product that might be relevant to any of the topics set forth above are welcome to include a description of that product. Comments about a specific technology or product should be limited to three pages or less.

As outlined throughout this response, Enveil's ZeroReveal® solutions leverage PETs to enable a decentralized approach to secure data collaboration, allowing participating entities to retain granular control and ownership of their sensitive data assets at all times — an approach that differentiates the company's solution from other Privacy Enhancing Technology-based offerings. Enveil ZeroReveal is a

lightweight, two-party, proxy-layer software system consisting of the ZeroReveal Client application and the ZeroReveal Server application. The Enveil Client application lives within the organization's trusted environment and is responsible for encrypting the operations/searches and decrypting the results. The Enveil Server application lives within the environment(s) of the data and is responsible for processing the encrypted operations over the data without ever decrypting anything.

By decoupling from the storage technology layer, Enveil is able to deploy above the data, allowing organizations to retain control of their data assets while leveraging existing storage methods and access controls mechanisms. Enveil focuses exclusively on solutioning to secure Data in Use and integrates with existing at-rest and in-transit data security solutions, acting in a complementary capacity to provide full lifecycle security via standard APIs. The deployment model is optimized for highly distributed, decentralized multi-party, and/or hybrid cloud deployments, such as the approach outlined in this RFI.

ZeroReveal is hardware agnostic and proven to work robustly and at speed over terabytes of data on cloud platforms and in third-party data locations. Further, Enveil's ZeroReveal software applications are built to integrate within existing systems. As such, the Client application is easily integrated with existing or new User Interface(s) that would enable a single intuitive, user-friendly web application for users to perform secure, federated searches and analytics across a variety of diverse datasets.

Enveil ZeroReveal® Product Suite (NIAP/CSfC Certified):

- **ZeroReveal Search** – Encrypted Search, Watchlists, Tipping and Alerting (version 4.x)
    - Core technologies used: FHE (Fully Homomorphic Encryption), SMPC (Secure Multiparty Computation), PSI (Private Set Intersection), PHE (Partially Homomorphic Encryption)
    - Modular encryption, leveraging open source FHE libraries

- **ZeroReveal Machine Learning** – Encrypted Evaluation (2020) and Encrypted Training (2022)
    - Core technologies used: FHE and SMPC (Secure Two-Party Computation)
    - Enables federated, encrypted machine learning
    - Modular encryption, leveraging open source FHE libraries

Overarching technical benefits of ZeroReveal:

- Decentralized framework means data stays where it is and owners retain control
- Completely API-based, users to interact with data in the same way that do today
- Sits above the data, requiring no changes to the underlying environment
- Works alongside existing at-rest and in-transit security protections

Proving that PETs are ready to be used at scale today, Enveil is deployed and operational today across the federal and commercial sectors. Here are four real-world examples of how Enveil's solutions are leveraging PETs to revolutionize data usage in the global marketplace today:

Cross-Jurisdictional Data Usage

Enveil worked with a large, EU-based financial institution to validate how ZeroReveal can be used to facilitate the secure and private data sharing needed to build a trusted collaboration network between entities. Enveil uniquely addressed specific customer pain points by enabling users to match customer profiles and enriched data across parties, as well as query for indicators and AML typologies across entities in order to offer additional insight on financial crime activity and behavior. Gaining access to a wider set of data in this secure and privacy-preserving manner improved outcomes by reducing false positives, driving prioritization, advancing the efficiency of financial crime investigations, improving enterprise data quality, and enabling greater operational efficiency.

As a two-party, proxy layer software solution, Enveil integrated with the enterprise's existing data structure and leveraged existing authentication, access control, and audit mechanisms. Data owners did not have to change their data environment or re-encrypt their data, notably shortening the time to value delivered by Enveil's solution. Enveil enabled encrypted queries across datasets of 100k and 1 million customer records. To further demonstrate scalability, a comparable third-party data source was used to attain and test datasets containing 5 million, 25 million, and 100 million records, respectively.

Enveil ZeroReveal allowed the large global financial institution to explore collaboration use cases and obtain access to additional data sources by providing strong security features, highly customizable business logic, and a fully traceable and transparent audit/regulatory control process. Enveil's PETs-powered capabilities delivered encrypted queries executed in the timeframe of (single digit) seconds to align with existing customer workflows. The engagement verified how Enveil's solution could be used to overcome legal and compliance boundaries by ensuring that sensitive data remained encrypted during processing and the security or ownership of the underlying data were never compromised. By leveraging a decentralized data model, Enveil ensured participants are never required to move or consolidate data assets, a requirement which is rarely feasible and has proven to be a barrier to success for similar efforts in the past.

Combatting Human Trafficking

Enveil partnered with DeliverFund, the leading counter-human trafficking intelligence organization, to harness their respective technology and data sourcing breakthroughs to advance initiatives for social good. The companies are working together to strengthen efforts to combat human trafficking, a growing criminal epidemic generating an estimated $975 million annually in the United States alone. With the largest analyst-curated human trafficking database in the U.S., DeliverFund has significantly reduced the time it takes to identify victims and those who exploit them, going beyond detection to work on the side

of prevention and risk mitigation. The collaboration with Enveil's PETs-based solutions expands the organization's impact by accelerating reach and efficiency.

Enveil's privacy-preserving ZeroReveal capabilities completely change the security paradigm for how users can leverage DeliverFund's counter human trafficking platform without the risk of exposing PII or other regulated data assets. Enveil's ZeroReveal search capabilities allow commercial organizations to securely cross-match and search DeliverFund's extensive database without ever revealing the contents of the search itself or compromising the security or ownership of the underlying data. By retaining control of their data at all times, organizations minimize their risk and ensure that sensitive and/or regulated data assets are never exposed to DeliverFund or any other entity. This enhanced privacy posture allows organizations to submit sensitive customer queries with confidence knowing that neither the data contained in the query or its corresponding results will ever be visible to DeliverFund or any other third party. By expanding the scope of trusted compute, organizations are able to extract value from data in untrusted spaces while remaining compliant with international privacy regulations.

Expanding Data Access at the Edge

Enveil teamed up with Terradepth, an advanced ocean data-as-a-service company revolutionizing ocean data use, to expand secure and private data usage and access at the Tactical Edge. This unmatched capability transforms the way ocean data can be leveraged for sensitive business and mission applications, including secure maritime domain awareness and mission planning.

Oceanographic data itself is a national and commercial asset and numerous industries, including oil & gas, national security, and telecommunications rely on its content and accuracy. Bringing together Enveil's ZeroReveal capabilities and Terradepth's oceanographic data holdings, the partnership allows organizations to access and utilize previously restricted datasets without revealing their interest or intent. Terradepth's cloud-based, market-leading ocean data platform, Absolute Ocean (AO), provides immersive and interactive visualization capabilities of both Terradepth-collected and third-party ocean data. Utilizing the integration with Enveil ZeroReveal, customers can leverage all these data holdings along with the full functionality of AO without revealing their interest and intent to Terradepth, Enveil, or any other entity.

Leveraging breakthroughs in PETs, Enveil's capabilities extend the boundary of trusted compute by securely processing data at the point of collection and in third-party data environments. A customer's sensitive search parameters, such as specific geographic areas of interest, specific objects, and data types, are protected through the encrypted search functionality powered by Enveil ZeroReveal. Organizations can maintain operational integrity during sensitive business and mission applications by ensuring the content of the search, watchlist, or analytics remain encrypted throughout the processing lifecycle.

Unlocking Secure and Private Third-Party Data Access

Enveil's partnership with Sayari, the global corporate data provider and commercial intelligence platform, increases data value by expanding secure and private usage and access. The technology and data integration provides customers in regulated industries access to leverage Sayari's extensive world-wide beneficial ownership and financial intelligence data without increasing organizational risk or requiring sensitive data to be moved or replicated. By enabling organizations to search Sayari's extensive data holdings without ever exposing their interests or intent, the partnership expands the ways sensitive and regulated data can be used. The collaboration unlocks value by enabling data to be used in ways that were not previously possible for sensitive business and mission applications.

ABOUT ENVEIL: Enveil is a pioneering Privacy Enhancing Technology company focused specifically on protecting Data in Use. Enveil's business-enabling and privacy-preserving capabilities change the paradigm of how and where organizations can leverage data to unlock value. Defining the transformative category of Privacy Enhancing Technologies (PETs), Enveil's award-winning ZeroReveal® solutions for secure data usage, collaboration, and monetization protect data while it's being used or processed. Customers can extract insights, cross-match, search, and analyze data assets at scale without ever revealing the content of the search itself, compromising the security or ownership of the underlying data, or exposing their interests and intent. A World Economic Forum Technology Pioneer founded by U.S. Intelligence Community alumni, Enveil is deployed and operational today, revolutionizing data usage in the global marketplace. Learn more at www.enveil.com.

December 22, 2022

To: White House Office of Science and Technology Policy (OSTP)
Subject: Key Digital Infrastructure for a Pandemic Preparedness and Response System for US Public Health

Vibrent Health
4114 Legato Rd. Suite 900
Fairfax, VA 22033

Please direct correspondence to:
Praduman Jain (PJ), CEO
pj@vibrenthealth.com

About Vibrent Health
Vibrent Health is a digital health, minority owned health IT small business headquartered in Virginia. Vibrent's mission is to accelerate health research to optimize human health through precision health technology and digital tools that empower participant engagement, research analytics, and research management for researchers, research organizations, public health professionals and research participants. We provide digital health infrastructure, apps and digital tools to government agencies, academic centers, health systems and life sciences companies for observational research, clinical trials and digital epidemiology.

The White House Office of Science and Technology Policy (OSTP), in partnership with the National Security Council (NSC), is leading efforts prepare for future pandemic-related national emergencies, disease outbreaks and other emergencies (both pre-emergency phase as well as emergency settings). This response to the RFI on "Data Collection for Emergency Clinical Trials and Interoperability Pilot" describes the Government's needs around viable strategies to distribute clinical trial protocols and capture clinical trial data using common application programming interfaces (APIs) and proposes approaches to address those needs.

The need for this RFI arises from the Nation's historical inability to react to exigent national health emergencies. This has largely been the result of a lack in intersectoral collaboration, and consequently, shortcomings in data harmonization, leading to inferior interoperability at scale. This shortcoming is all too common in large-scale health surveillance and clinical research, particularly given the solutions already developed by the National Institutes of Health and partners. It is this investment by government, academic, and industry experts that nationwide emergency clinical trial infrastructure should be predicated upon. This is temporally, economically, and scientifically the most efficient and sensical process.

Quickly and widely distributing clinical trial protocols and equitably capturing tangential or resultant data at a nationwide scale for public health emergencies will require seamless interdisciplinary and intersectoral partnerships. This RFI appropriately identifies health IT companies as core to both informing and operationalizing an architectural solution to this challenge. Health IT companies, and specifically digital health technology companies, are ideally positioned to modulate the exchange of information (and material) in any complex health network, precisely because they are most fundamental to data structuring and traffic between interfacing partners (Jain and

Klein, 2019). Digital health companies are at the vanguard of addressing lacunae in the national data collection and clinical trial protocols. This RFI response describes the approach that a digital health company (Vibrent Health, or "Vibrent") has taken for a large, distributed nationwide data collection and data harmonization research initiative. Vibrent has been the technology backbone of the National Institute of Health's (NIH) nationwide All of Us Research Program since inception. Vibrent provides all digital infrastructure and data analytics NIH to enroll over one million individuals to study the effects of biology, lifestyle, and the environment affect human health. This cornerstone of Dr. Francis Collins' legacy has resulted in the electronic informed consent, data collection, data harmonization and return of results of genomics, phenotypic, EHR and demographic data of over 570,000 people using APIs, HL7 and FHIR protocols (All of Us Research Program, 2019).

Best Practices for Electronic Informed Consent at Nationwide Scale in Clinical Trials with Diversity Inclusion

APIs designed e-consent digital health clinical trials can be effectively deployed and managed to inform potential study participants of any risks and or benefits directly, and to collect electronic informed consent forms that can then be directly uploaded to a patient-accessible platform. Potential participants can be directed to download an API on any internet-accessible device or through QR codes placed strategically by outreach campaigns, healthcare provider offices, and public health entities. They should also be able to directly access the platform on any internet-accessible device. E-consent forms can be signed and dated electronically and validated with the co-uploading of government issued ID, and if necessary, a photo of the said individual then holding the government issued ID (as is common in transnational legal proceedings). Validation can also occur through SMS or email validation and other means. Forms can be made available in various extension formats, for user ease. Users can then transition between various translations of the forms as well as interact with disability-informed designs. Using such a system, version control can be easily incorporated, with digital time stamps, and proper documentation automatic. Also, regulatory adherence across jurisdictions can be ensured.

The e-consent process itself can become much more interactive for the user, with elegant UX design solutions. For example, supplemental information can be offered to the potential participant, if requested, or upon "interpreting" the user's needs. Software should be able to relay the individual to an FAQ page or an automated system that can address any concerns regarding health privacy rights and legislation. The API can incorporate rich media, such as advanced infographics and videos, accommodating diverse learning styles, reducing user effort (and hence completion rates) and providing clarity. If, after launch any refinement needs to be made to an e-consent form, or if, as the public emergency itself mutates, it should be capable of being amended and deployed in real time. Software should track which e-consent iteration the participant consented to. Software should also be able to re-send e-consent forms to participants, or reminders, if necessary. Evidence suggests that this level of bidirectional interactivity keeps the participants more engaged and willing to participate.

Consent forms and PHI should and can be appropriately organized and secured, as well as centrally stored on a cloud-based repository, managed by an assigned group of digital health IT administrators that have overlapping access to secure data keys. These

administrators will be responsible for granting access to a wide range of researchers nationwide, who can benefit from the use of a given participants health data. Policy should determine the process for granting access to the said data. Instead of an iterative process that begins with the individual researcher requesting data, data should be sent to researchers for use based on a previously agreed upon algorithm. Researchers will have completed forms stating their research interests and background (with credentials validated). Therefore, when in an emergency situation, data can be "pushed" directly to the researchers based on their data needs. Also, the platform should be able to "push" data (depending on health privacy legislation) to institutions or departments, to then be disseminated to individual researchers. This system will require a robust policy infrastructure, as well as a centralized registry of nationwide participant researchers, and institutions. Hubs should be determined, in advance, based on cascading geography and sector (region, state, institute).

E-consent forms should be coded according to random alphanumeric generators. Policy should be developed by digital health security experts. The receipt and centralized storage of e-consent forms and other PHI (protected health information) can be managed by a centralized hub.

Consent forms should be designed to maximize the re-usability and applicability of the data acquired. Consent form language should be crafted with the assistance of clinical trial IRBs such as that already oversee multisite clinical trials (such as University of California BRAID, or the State University of New York). Consent should ask for data that is of is of ongoing longitudinal value, and while collected during a limited period (presumably an emergency period), should be able to be queried periodically, into the future (after permanent de-identification). Using a purpose-designed participant management API, consent forms can offer multiple levels of consent, allowing the user to choose what they feel most comfortable with, therefore excluding as few individuals as possible, and fostering trial diversity. Digital health platforms can structure diverse data types for longitudinal collection or use, in a way that matrices multiple data types, such as phenotypic, clinical, environmental, sociocultural, demographic, or molecular inputs. This is critical to allowing researchers to ask unique and novel scientific questions, and for the collected data to provide the best long-term value for the public health of the nation.

Best Practices for Data Collection, Data Sharing and Workflow Automation

It is imperative that any API or platform allows for bidirectional communication between the study participant and research administrator (or researcher), in multiple formats. To cultivate engagement and an interest in altruistic science, the return of general and bespoke results to participants and participant communities is paramount. Studies show that the return of results foster participants that value an ongoing scientific relationship. It is also part of an ethical, trust-building exchange (Plunk et al., 2022). Results can be disseminated in very pointed means, to individual participants or through more general means, with the recognition of a community-based participatory research approach. Either way, this is the foundation of trust that is necessary between the scientific community (and government) and the broader American populace, to not only quantify and design diagnostics and therapeutics, but to have them adopted.

Of course, a participant should be able to cease participation in research at any time, without any stated reason. An API or platform should allow a participant to easily request that participation stop immediately or at a given time, which will automatically remove them from any study, send notification to any researcher or administrator for whom data was sent, and begin the process of data deletion. It should also ask the former participant if they would be interested in answering any questions regarding their desire to withdraw, so that future communications and study design can be optimized. The process should take into account standing legislation, such as HIPAA, NIST, FISMA, FIPS, SOC 1 and 2 Type II, PCI DSS, and FedRAMP.

Ideally, an API or platform would be designed for the purposes of the sponsor (researcher/administrator/healthcare provider). Such platforms are in addition to a platform that has been purposefully designed for the research participant. An ideal API would possess breadth and plasticity and be able to be nimbly and quickly refined by a software programming team. It would be able to streamline research processes from the sponsors perspective and assist with the organization of downstream data biosamples.

It is crucial that given the diversity of potential users, there is built-in flexibility to any API offering multiple methods of completing milestones, relevant to different user types. The API should most definitely be capable of supporting both digital and point-of-care data collection, scheduling and survey response collection. It should be at its core, purpose-built to streamline and simplify communication between those conducting clinical trials and participants. It should also, however, be designed for communication between researchers, administrators, and health providers. In a single location, researchers and administrators and particularly providers, should be able to efficiently manage all engagement, virtual/in-person appointments, survey responses, EMR abstractions/cases (and eCRFs), and communications with the participant.

Staff should be able to access and easily navigate this platform to ensure seamlessness with participant communication and follow-up, through digital, telephone, SMS, direct mail, or face-to-face interactions. For both scientific validity and medical ethics, it is vital that communication methods between those conducting research and participants is broad. Overly-complex systems can preclude those without digital comfort or even without access, i.e., the “digital divide”. eCRFs can most definitely be sent through a common health IT platform to administrators at the sponsor institution. From there, they can be filtered from the administrator to individual researchers and providers through the institution’s IT system. Furthermore, communications can be sent directly to the researcher or provider if they are granted access. Recommendation is that both institutions and individual end providers/researchers both have access to the common API, although through slightly differentiated user interfaces.

Any clinical trial, for scientific validity, should meet specific diversity goals (Ramirez et al., 2022). Real-time dashboards and reports should be available either episodically or upon request. This will supply the provider, administrator, and/or researcher with any requisite clinical or other health-related data in a straightforward manner (Schilling et al., under revision). The user should, moreover, be able to pull consent data or demographic data, allaying the time requirement for assessing compliance.

Survey completion rates can be increased by using computer-assisted telephone interviews (CATI) with those that have reduced digital access or comfort. Surveys can

also be deployed and indicated electronically through an automated system that integrates with the USPS or other mail carriers. An API designed for the researcher/administrator/provider should allow individuals to customize a completely configurable outreach website. One can capture information about prospective participants to develop into a registry, then run automated or high touch enrollment campaigns to turn prospects into consented participants.

Achieving recruitment targets can become simplified with the ability to create custom engagement campaigns that educate and encourage enrollment. Specifically, call lists, email campaigns, SMS campaigns, or direct mail can be easily facilitated. Dashboards should then analyze the success of various modes of outreach and communication, so that messaging can be refined.

It is quality UX design that makes users more likely to participate and complete fully, whether clinical trial participants or researchers/providers/administrators. Generally, any request for data should be ordered appropriately, worded simply, and flexible. Certain response types should be mandated/validated (e.g., no age over 130 years should be reported) however, input forms should reflexively allow the user to customize answers too, should they choose. This shows respect for the users input. Although the more data collected, the more robust studies will be, no data should be collected unless it is entirely necessary. It is therefore vital that EMR/EHR systems are full integrated, and that sample/specimen processing is largely automated, with logistics experts and "service experts" at the ready for any logistical hang-ups. An API for data analytics collect data in quantitative or qualitative form, when possible, rather than categorical data. This will allow the researcher to convert to categorical data, if necessary, but not preclude him/her from the source data, which can drive the greatest statistical power.

Digital methods have been widely shown to allay operational constraints and increase biomedical and healthcare efficiencies (Bombard et al., 2022). With any clinical trial specifically, leveraging digital systems to assuage the vagaries of non-digital processes, as well as data analytics increases the likelihood of both the provider and participant's full engagement. This includes the collection and storage (biorepository) of any necessary biological sample or specimen, or wearable, and the resultant data produced. An ideal API will be able to automate the scheduling, fulfillment, and tracking of any sample/specimen collection kit to the participants residence, processing laboratory, and long-term, secure repository (Jain, 2021). An apropos example is for genomic testing panels. This will save time for both the participant and the provider. If a procedure needs to be conducted on-site, the API should be able to connect to a network of partner organizations where this can occur (for example, for urinalysis, blood draws, etc.). The integration of the digital and the non-digital is a real estate that should be fully developed for user ease, adoption, and adherence, whether study participant or study conductor. Processes such as artificial intelligence and machine learning, that is ethically and expertly designed can assist with automatic triaging and workflow automation (Schilling et al., in press)

TEFCA SOPs should consider communication pathways between clinical trial sponsors and local (city and county) and state health departments. Currently, there is limited formal communication between health departments and the healthcare systems that exist in their jurisdiction. Any formal communication and downstream agreements are

generally piecemeal, the product of various contracts that had specific goals in mind, rather than the creation of a multi-use communication floor. Consequently, there is a lack of patient record tracking across healthcare centers, or the provisioning of wraparound services. Health departments are well positioned to liaise between community organizations, healthcare centers, and government to facilitate emergency clinical trials. There is, however, scant ability to tailor trials to local health conditions on the ground due to a lack of effective overarching communication infrastructure. Ideally, healthcare centers should have a streamlined way of communicating hospital and provider metrics (bed occupancy, etc.) to public health entities. Health departments should be able to provide relevant epidemiological reports to healthcare centers, customized to their geography or service population. Although there are great examples of collaborative efforts, a common API that can be accessed by both, specifically in the context of conducting emergency clinical trials will allow for much greater interoperability at smaller jurisdiction levels (e.g., city, county, and state) (Alemi et al., 2022).

The benefit of a participant-centric, digital-first approach at-scale is that it emphasizes systems biology and implementation science. This honors the complex design of nationwide clinical trials, that often require sociocultural, environmental, medical, and/or phenotypic data to effectively, ethically, and quickly design and manufacture prophylactics, diagnostics, and therapeutics (Ginsburg et al., 2021). Because we “do not know what we do not know” it is critical to depend on a system that has the plasticity to incorporate diverse data streams. It also fosters a longitudinal approach, making it easier to consider the effects of shifts in environment, epidemiology, policy, or molecular evolution of an infectious pathogen (affecting virulence, resistance, transmissibility, etc.)

Significance

Much of the investment in creating a digital infrastructure for distributing clinical trial protocols nationwide, capturing clinical trial data, and more has been completed by the National Institutes of Health, effectively creating a blueprint. Digital health companies such as Vibrent are well-positioned to collaborate with multiple intersectoral partners, overseeing vast communication pathways. We can avoid costly mistakes and reduce risk by following an established track record.**Recommended Approach for a 6-12 Month Pilot Project**

Critical Digital Health Infrastructure for Pandemic Preparedness and Response for the US Public Health on a National, State and Community Level

The work being proposed in this pilot is an extension of the research and development of the Vibrent digital health technology infrastructure which **has already been funded by the NIH All of Us Research Program, RADx Tech and RADx UP.**

This research and development government funding to Vibrent for this clinical trials digital infrastructure creates a blueprint for the kind of digital epidemiology technical infrastructure needed for future pandemic preparedness. Vibrent is well-positioned to collaborate with multiple intersectoral partners while overseeing national and local communication pathways. The government can avoid costly mistakes and reduce risk by following an established track record and using what has already been funded.

The US Government needs a method of rapidly reacting to a public health emergency by distributing clinical trial protocols and capturing clinical trial data using common

application programming interfaces (APIs). The Government would benefit from a pilot project that stands up a reference implementation of the technology platform utilized for the NIH's RADx TECH and RADx UP progarms.

This approach describes the value of a 6-12 month pilot project to demonstrate data standardization, data capture, data harmonization and data sharing for a national digital public health infrastructure. A "**Public Health Emergency Clinical Trials Platform"** implementation pilot project would be 12-months in duration and:

1. Implements the platform within 60 days,
2. Loads a base protocol into the platform,
3. Demonstrates how the protocol(s) will be distributed to clinical trial sponsors SMART on FHIR APIs,
4. Collects contact information, demographic data and EHR data from thousands of consented permanent residents of the US,
5. Demonstrates how the collected data will be made available to clinical trial sponsors, public health leaders and other public health stakeholders and
6. Demonstrates how the registry cohort can be maintained on an on-going basis and used as an on-demand turn-key mission-critical registry for public health emergencies, ready at a moment's notice for clinical trials.

At the end of this pilot project, the US Government will have a public health digital clinical trials digital epidemiology infrastructure based on a federally funded and scientifically validated technical platform for clinical research. The digital tools have been validated with diverse populations and communities that are typically under-represented in clinical research.

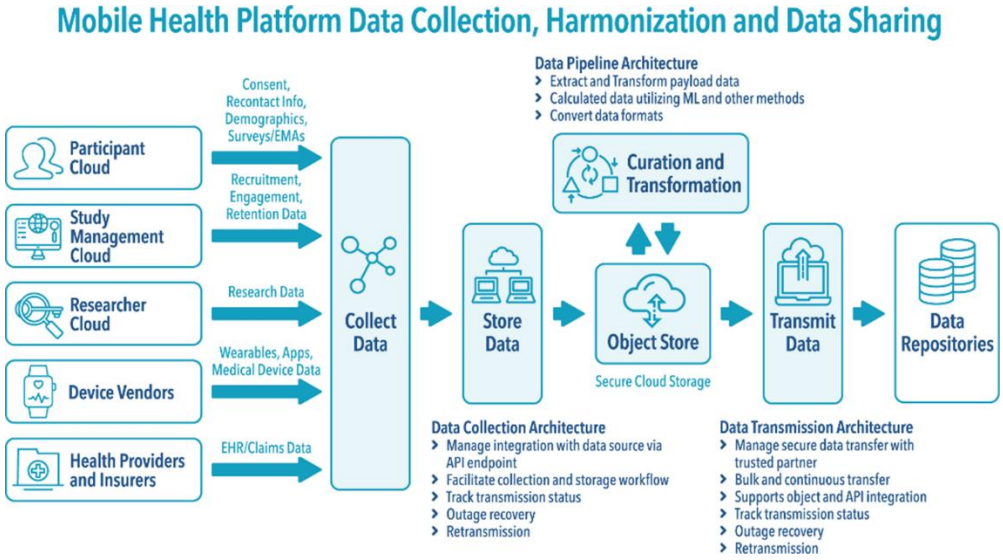
Benefits for the US Government and US Population

There are many benefits of this Cloud-based, cybersecure reference implementation pilot project for the US Government and the US population. The pilot project would provide national and local public health departments with a cybersecure digital health technology infrastructure for public health can both be used as a public health emergency digital cohort in non-emergency times and as a turn-key clinical trials registry during pre-emergency and emergency times.

The digital infrastructure will be readily available for rapidly loading many different concurrent e-consent and data collection protocols into the platform, ready to collect, share and distribute data to clinical trial sites and authenticated public health professionals in real-time. The digital infrastructure includes the technology tools necessary for decentralized, national recruitment, engagement of both the underlying public health registry cohort as well as an unlimited number of individual clinical trials. The infrastructure includes core components of cybersecurity, account management, protocol management, FHIR Questionnaires, SMART on FHIR APIs, EHR APIs, data integration APIs, data harmonization, low-touch digital first recruitment and engagement tools and high-touch digitally-enabled recruitment and engagement tools.

This adaptable, decentralized technology platform approach allows the US Government to anticipate that which can be anticipated but remain flexible enough to reasonably accommodate the unexpected prior to and during public health emergencies.

Figure 1: The proposed pilot project will utilize a Cloud-based, cybersecure digital health infrastructure validated at scale by the NIH



The Digital Clinical Trials Infrastructure for the Proposed Pilot is Validated by multiple nationwide NIH Mission-Critical Contracts

This digital health infrastructure has been validated by the NIH with over 570,000 consented participants, 300,000 Covid survey responses, 5GB worth of Fitbit data, 630 GB of Apple Health Kit data (health data, EHRs), 10,415 at-home genomics salivary kits delivered and processed and thousands of professional end-users at clinical trial sites nationwide. The digital health infrastructure has ethically returned COVID serology results direct to 24,000 participants.

The digital infrastructure has also been validated via a $4.23 million RADx contract with the National Institutes of Health for providing a digital infrastructure and AI/ML for pandemic testing and control called “Digital Health Solutions for COVID-19”. The goal of this initiative was to support the development of digital health tools that could leverage multiple data sources, privacy-preserving technologies, and computational tools to assist in managing population health during the COVID-19 pandemic. This Vibrent project has been funded with Federal funds under Contract No. 75N91020C00038.

The RADx contract was to build and validate an NIH-funded AI/ML symptom screening tool and integrated it with the already robust digital health infrastructure to make at-home testing more accurate, convenient and adaptable to rapid changes in outbreaks, epidemics and pandemics. The RADx clinical trial validated the usability and acceptability of the technology infrastructure for collecting data, informing participants of their COVID-19 risks, and accessing possible infection with SARS-CoV-2 through pilot testing in African American participants (J.Schilling et al., in press, under revision).

Through the RADx contract Vibrent has developed and scientifically validated a digital epidemiology public health tool powered by artificial intelligence (called “COVID-CARE”) which provides for infectious disease screening, testing, and contact tracing through a mobile application. Vibrent’s technology can improve disease containment efforts through its precision screening capabilities that can ensure that individuals are tested in a timely manner. Ensuring that testing is both targeted and meaningful is crucial to

delivering treatment to those who need it most while also providing the most impactful data for how a particular infectious disease spreads. Vibrent's screening tool indicates when a test is medically needed, which promotes not only accurate disease detection but also accurate decision-making about next steps in care. This technology provides a model for a new wave of health technologies that can change the dynamic of pandemic response while conserving resources and health expenditures.

Vibrent's use of AI/ML is informing a new line of thinking related to combatting infectious disease, which involves not only the immediate testing and treatment of individuals for COVID but also the collection and analysis of data on a macro-scale to inform new public and private protocols for containing future pandemics. This understanding will help our country more quickly stop the spread of a virus through the establishment of herd immunity at the earliest possible date.

The platform incorporates participant web and mobile apps along with interactive dashboards and tools to provide real-time analysis of incoming data to inform future public health interventions. This technology is highly cybersecure, with Vibrent receiving an Authority to Operate security certification from the NIH in 2017.

From the NIH RADx Program Officer for the Vibrent contract:

> *"The solutions developed by Vibrent Health and the other project teams have clearly demonstrated that digital health tools will play a crucial role in enhancing the use of various physical technologies going forward. In the future, public health officials, clinicians, employers, and the general community will rely heavily on these digital tools as they work to provide individuals with the necessary data to make informed choices about their health."*

Versatility of a Digital Health Infrastructure for Human Subjects Research

The versatility of a cloud-based, multi-modal approach means that bidirectional communication between researcher/administrator/provider and participant (or sponsor IT department) can be achieved in real time, improving survey response rates, scheduling, data collection and scientific collaboration. Participant experience platforms recognize the co-constitutive nature of modern clinical trials and are ideal for managing a hassle-free experience for all users.

Digital methods have been shown by Vibrent's collaborations with scientific partners to reliably increase study diversity. Regarding the All of Us research program, using the Vibrent platform (Research Cloud, Participant Experience Manager, and Data Explorer) >570,000 individuals have been e-consented thus far (~80% of which are underrepresented in biomedical research and ~50% of which are ethnoracial minorities), 390,000 biological samples have been collected, and >322,000 EHR records have been incorporated. More recently, health data from ~20,000 COVID-19 patients, and >57,000 responses to SDoH surveying have been made accessible to researchers.

Other large scale clinical research studies have also been built using Vibrent Health's digital platform as a foundation including:

- The "Mount Sinai Million" genomics precision health study (Mount Sinai Health System in New York) which has utilized experience management and research cloud operability to begin enrolling 1 million individuals across 8 clinical sites, each with linked genomic-phenotypic data and

- The My Healthy Maryland statewide genomics epidemiology study which will enroll 250,000 participants over a 5-10-year period, and integrate EHR data from 14 healthcare *systems*, as well as extensive biological sample and environmental data, with sophisticated informatics and artificial intelligence algorithms.

Vibrent looks forward to serving as a partner to the OSTP on these important public health preparedness efforts to ensure a coordinated and informed response to future biological threats that improves the health and safety of the public.

Citations

Bombard Y, Ginsburg GS, Sturm AC, et al. 2022. Digital health-enabled genomics: Opportunities and challenges. *American Journal of Human Genetics*, 109:7, 1190-1198.

Ginsburg GS, Penny M, Feero WG, et al. 2021. The National Academies' Roundtable on Genomics and Precision Health: Where we have been and where we are heading. *American Journal of Human Genetics*, 108:10, 181701822.

P. Jain. "Could Biobanks Be the New Frontier for Digital Health Research? Vibrent Health argues that ongoing interactions between biobanks, researchers, and biospecimen donors can build trust and reinforce precision medicine." Genetic Engineering & Biotechnology News 41.11 (2021): 22-23. https://doi.org/10.1089/gen.41.11.08 Link to paper

Alemi, F., Vang, J., Wojtusiak, J., Guralnik, E., Peterson, R., Roess, A., & Jain, P. (2022). Differential diagnosis of COVID-19 and influenza. PLOS Global Public Health, 2(7), e0000221.https://doi.org/10.1371/journal.pgph.0000221 Link to paper

Plunk, A., Sheehan, B., Orr, S., Gartner, D., Moeller, F., & Jain, P. (2022). Stemming the Tide of Distrust: A Mixed-Methods Study of Vaccine Hesitancy. *Journal of Clinical and Translational Science,* 1-12. doi:10.1017/cts.2022.492 Link to paper

Ramirez, Andrea H., et al. "The All of Us Research Program: data quality, utility, and diversity." Patterns 3.8 (2022): 100570. Link to paper

J. Schilling, F.G. Moeller, R. Peterson, B. Beltz, D. Joshi, D. Gartner, J. Vang, P. Jain. "Testing the Acceptability and Usability of an AI-Enabled COVID-19 Diagnostic Tool among Diverse Adult Populations in the US." In Press

J. Schilling, D. Klein, M.M. Bartholmae, S. Shokouhi, A. J. Toepp, A.A. Roess, J.M. Sill, M.V. Karpov, K. Maney, K. Pearson Brown, B.L. Levy, K.D. Renshaw, S. Dodani, P. Jain. "COVIDsmart: A Digital Health Initiative for Remote Data Collection and Study of COVID's Impact on the State of Virginia". Under revision

Shah RN, Nghiem J, Ranney, ML. 2021. The Rise of Digital Health and Innovation Centers at Academic Medical Centers: Time for a New Industry Relationship Paradigm. *JAMA Health Forum*, 2:3, e210339.

All of Us Research Program Investigators et al. The "All of Us" Research Program, New England Journal of Medicine (2019). Link to paper

RFI; Data Collection for Emergency Clinical Trials and Interoperability Pilot

Prepared by: Gav Martell, YonaLink Inc
gav@yonalink.com
416-662-5074

December 27, 2022

Gav Martell, the respondent, is co-founder and VP of Business Development of YonaLink Inc, a company established in the state of Delaware, with offices in Boston, MA. YonaLink is a company that provides software as a service for clinical trials. Specifically:

1) As a platform to stream data from Medical Center EHRs (Electronic Health Records) and other eSources to the trial EDC (Electronic Data Capture system)
2) Provides a next generation EDC system that was built specifically with data streaming in mind, including all the workflow and tools necessary to stream data from multiple sources into the EDC

YonaLink's FHIR-based system is custom built specifically to leverage the technology mentioned in this RFI, and uniquely has a global solution already in the marketplace.

Comment:

The solution outlined in this RFI in terms of a data capture system is already in place - which allows for coordinated and large-scale clinical trials to be efficiently carried out across a range of institutions and sites as needed to address outbreaks of disease and other emergencies. YonaLink is a Saas system that allows for the distribution of clinical trial protocols and capture of clinical trial data using common FHIR-based APIs. Case report forms can easily be built and disseminated among a network of sites without custom installation or integration at the site level. In addition, YonaLink can today stream data from EHR systems at tens of thousnads of US Medical Centers to the data collection system. The system is operational in the U.S., and a pilot could be started in a matter of days or weeks with little effort. One of the advantages of this solution is the reduced burden on site staff, who no longer need to spend time manually copying data points from their EHR into the EDC. This reduces staff burn out, increases productivity, reduces timelines, and ultimately lowers the threshold for smaller sites to participate in clinical trials. In doing so, the pool of sites available to participate in clinical trials is expanding, including sites in underserved communities, thereby enlarging and strengthening the overall clinical trials infrastructure.

For the desired use case, in step 4, rather than send the eCRFs via common APIs, we believe it would be more effective to have the eCRFs available in the same digital system that will transfer the data to the EDC. That is to say, traditionally, data is manually entered into an EDC, and verified before being made available to the trial sponsor. In a system such as YonaLink that is built to stream data, there must still be a verification process that allows the study coordinator to verify that the data being streamed is accurate, before it is made available to the sponsor. This requires them to access such a system on a routine basis. As such, it would make

sense for this system to be the same one that houses the eCRFs. Rather than having sites or a vendor build a system that can receive eCRF APIs, it would make sense to leverage the EHR-to-EDC system as the one point of login and use.

YonaLink has two levels of interoperability, both at the patient level, as well as at the site level. By leveraging the 21st Century Cures Act, as well as SMART on FHIR technologies, YonaLink has a global solution that functions across different institutions and EHR systems, and also provides adequate functionality to support emergency clinical trial research.

In general, studies with more structured data are better suited for obvious reasons. We would be happy to discuss de-identification, and other functionality already in place in our system.

A very effective pilot could be executed with the following:

- a trial sponsor
- an EHR-to-EDC vendor such as YonaLink
- a data analysis vendor
- participating sites

An end-to-end solution that allows for the emergency preparedness outlined in this RFI is not future-ware, but is prepared today to run a pilot, and be scalable on a national or global scale.

TO: Jennifer Roberts, Assistant Director for Health Technologies
White House Office of Science and Technology Policy, and
Grail Sipes, Assistant Director for Biomedical Regulatory Policy
White House Office of Science and Technology Policy

FROM: Teri Marlene Prince
CEO / President, TERIDA

DATE: December 27, 2022

RE: **Response to the Notices in the Federal Register, and specifically, to the following needs:**

- a coordinated clinical trials enterprise that can swiftly characterize emerging viral threats and evaluate the effectiveness of vaccines, therapeutics, and other countermeasures across a diversity of trial participants
- a U.S. clinical trials infrastructure "ready to administer candidate countermeasures to participants within 14 days after the identification of a viable countermeasure."
- a coordinated clinical research system can be deployed in the event of an emerging disease outbreak efficiently, effectively, and securely
- a technical solution that ensures that trial data can be captured as a set of consistent data elements across separate trial sites under a coordinated clinical trial protocol(s) with secure data repositories, accessible to those with the appropriate level of permissions for rapid pandemic preparedness and biodefense.

Why am I writing at this last moment? The reasons:

- today's email,
- the legislative framework for cybersecurity now codified explicitly with the FedRAMP Act inside the 2023 Defense Authorization Act passed, and signed by the President, and
- TERIDA able to disclose that the United States Senate is taking TERIDA / the enterprise version of the *Terida RegTech Framework – CLASsoft*™ cloud platform to FedRAMP ATO.
    - See attached Exhibit for TERIDA's RegTech cloud platform (12/3) listing on FedRAMP.gov, and
    - the link below for the FedRAMP PMO approved (12/22) press release about these cloud endeavors
      https://www.einpresswire.com/article/607731019/terida-achieves-fedramp-in-process-designation-for-its-regtech-framework-classoft

To summarize: Respectfully, we propose the above innovative TERIDA RegTech cloud platform for configuration / customization / extension to meet the above objectives within the demonstration timeline.

The following Service Description of the TERIDA RegTech cloud platform (copied from FedRAMP.gov) summarizes our enterprise solution and wide-ranging use cases:

The Terida RegTech Framework – CLASsoft™:

One Framework, Infinite Applications.

Robust, scalable, e-operations RegTech platform to receive, manage, track, monitor, analyze, evaluate, resolve, and audit registrations, applications, users, communications, claims, cases, files, forms, and documents.

With CLASsoft, all information collected and processed, and their chain of custody and access, are secured, managed, and protected for the term(s) necessitated by operational requirements and objectives, and rules, regulations, laws, and evidentiary, audit and risk standards.

Credential and connect the enterprise with the Terida RegTech platform's consistent methodologies and permission layers. Reduce data silos. Decrease costs and risks.

With CLASsoft, the entire flow of information "Persons, Objects, Events, Organizations, Entities, Registrations, Applications, Claims, Cases, Files, Data, Documents, Forms, Communications, Relationships, Requirements, Regulations, Operations, Processes, Protocols, Workflow, Permission Hierarchies, Administration, Evaluation, Resolution, Reporting, Auditing, Archiving" is configured explicitly to the particular business problem and secured within the platform's authorization boundary.

E-credential, E-submit, E-consent, E-certify, E-participate, E-claim, E-evaluate, E-process, E-learn, E-work, E-connect use cases include:

- government, business, organizations, defense
- education, finance, health, insurance, legal
- emergencies, disasters, evacuations, mass actions, class actions
- device failures, complaints, credentialing, compliance
- claims, settlements, distributions, deferred prosecution agreements
- joint operations, collaborative intelligence, procurement, supply chain.

In terms of the RFI and ongoing ONC objectives: please contact us to discuss our wide-ranging secure cloud solution capabilities, and our personal, corporate, and teaming experience.

Establishing the necessary platform for secure and responsible sharing of information between health networks for emergencies, pandemics and defense requires more than interoperability of records and legacy systems. It requires an understanding of the past, the present and the future.

With more than 50 years of academic, corporate, consulting, and personal IT interoperability experience, I would very much like to be in touch.

CORPORATE SUMMARY

TERIDA LLC **40 Augusta National Drive** **Pinehurst NC 28374** **https://www.terida.com**	**DUNS:** 134500599 CAGE: 3YDU2 **SAM UEI #:** F95THAVKNXP8 **DD2345 JCP-DLA CERTIFIED?** YES - CERT #: 0085194 **SMALL BUSINESS?** YES – WOSB and EDWOSB
POINT OF CONTACT	Teri Marlene Prince Terida President / CEO 910.692.4678 (Office) / 910.603.0104 (Mobile) tprince@terida.com ; RegTech@terida.com

OUR WORK / OUR CREDENTIALS:

- award-winning, highly secure, enterprise, multi-use case, RegTech Platform cloud solutions
- US, Canada, EU compliance, cybersecurity, privacy (NIST, FISMA, FedRAMP, StateRAMP, CMMC, CSPV-SEE, GDPR)
- Efficiency, Effectiveness, Experience, Agility, Innovation, Cybersecurity
- Consulting, Single Language / Multi-Language RegTech Platform Deployments - complex, but user friendly, hundreds of thousands of users, distributions totaling more than a billion dollars, extensive long term case load
- Non-Traditional Defense Contractor - OTA Consortiums Memberships: currently, TERIDA is a member of five DoD Other Transaction Authority Consortiums:
  National Security Technology Accelerator Consortium (NSTXL), Space Enterprise Consortium (SpEC), Resilient Infrastructure + Secure Energy Consortium (RIS), Senior Healthcare Innovation Consortium (SHIC), and Medical Technology Enterprise Consortium (MTEC).

NATO COMMUNICATION AND INFORMATION AGENCY – BASIC ORDERING AGREEMENT – NCIA BOA #14369:

- TERIDA CLASsoft™ RegTech cloud platform contract vehicle available to all NATO bodies, all NATO nation government agencies, military forces and contractors, and other nation/forces/contractors with NCIA approval.

FedRAMP / StateRAMP – UPDATED STATUS

- With FedRAMP Readiness (RAR 2018 and RAR 2020) and the United States Senate contract award, TERIDA moved to 'FedRAMP In Process' status (United States Senate, FedRAMP SAR / ATO authorizing "agency"), November 2022
- TERIDA named to StateRAMP's product list based on FedRAMP credentials. Our StateRAMP sponsor – the North Carolina Military Business Center (NBMC: NCMBC.us, DEFTECH.nc.gov, MatchForce.org, CyberNC.US). Currently, TERIDA is one of the very few StateRAMP Product List vendors with a StateRAMP Sponsor/

From a technical perspective, OSTP is also seeking input on how best to operationalize both protocol distribution and data capture in a forthcoming RFI.

Information Requested: Respondents may provide information for one or as many topics as listed as they choose.

Introductory Comments:

"I am submitting input on the sections described following this introductory paragraph. In short, my input is based on the application of a blockchain technological approach to research. I am the president of Suncoast RHIO, Inc. (Regional Health Information Organization) based in Florida and doing business nationwide. We have a USPTO patent pending submission for "The Healthcare Blockchain" and a trademark for the term as well.

We applaud the efforts of the ONC as it dovetails with our interoperability approach its applicability to this research RFI. We know the approach can be inherited to our target blockchain solution in support of this RFI. We are, in practice, utilizing ONC's work in establishing the regulatory and governance foundation for the interoperability of electronic health records, healthcare technology, and in the highest level of the overall model template of the practice of medicine reflected in technology for our nation. Our company and our efforts in blockchain are focused on capturing the granular data elements of healthcare by utilizing a dynamic designated record set model.

The ONC's work in the development of the United States Core Data for Interoperability (USCDI) standard; the FHIR APIs; and Substitutable Medical Applications and Reusable Technologies (SMART) platforms that are compatible with FHIR interfaces, parallel our efforts in using our efforts listed including "SMART on FHIR" with our own open and CURES certified EMR and others. We utilize APIs and UDS+, to map FQHC data granular data elements with our Interop X software for FHIR mapping and eSignature just to name a few. It is with this background that we submit our responses to the RFI focusing on a hybrid blockchain and traditional platform."

1. *Governance for emergency clinical trials response.*

"Our reading of the questions and topics listed in the RFI that are of procedures and process orientation will not be addressed by us. These areas are unique in our view to those familiar with research and research related public outreach. We are focusing primarily on three sections to respond; The first is this section #1 on all subsets of Governance; the second is on Section #2 regarding community-based care networks, and the third on Section #4 regarding confidentiality."

a. **RFI-** Descriptions of models that could be used to establish a U.S.-level governance structure for emergency clinical trials. As noted above, one possible approach would be a centralized U.S.-level structure drawing membership from Federal agencies with relevant expertise.

COMMENT- "As stated, we understand that a key component in building U.S. capacity for clinical research is ensuring that trial data can be captured as a set of consistent data elements across separate trial sites under a coordinated clinical trial protocol. We add to this subset of metadata of data elements to be the administrative granularity of patient information beyond clinical elements. By using a blockchain ledger type approach, items such as consent and electronic signature for example, in all of their varied and legal forms, are capable of being captured and protected from misuse by preservation of their legal uniqueness. A method of mixing hybrid NFT and fungible data elements could be employed.

As stated by the Advisory Group Company at the Blockchain in Health Care Educational Briefing for Non-IT Executives International Global eHealth Executive Council Executive Summary, Blockchain is a digital ledger that enables parties with no history of knowing or of having a history of working with one another to securely commit to multi-entity transactions without the need for an intermediary."

1. d and e **RFI-** Communications and Tracking

COMMENT- "Blockchain technology allows decentralized communities of people and organizations to establish a single, shared record of events with confidence that no one can tamper with historical records once they have been verified.

With a shared system of record that either every person or select provisioned persons in a blockchain is guaranteed to see as defined by policy, the same historical record is accessible. It is tamper resistant in that once data has been recorded on the blockchain, it is effectively permanent."

1. i. **RFI-** Optimal ways to manage interactions with domestic and international regulatory bodies.

COMMENT- "Blockchain is a global application is used in most countries in Web1, Web2, and the upcoming distributed network database based Web3. It is not a currency."

1. j. **RFI-** Appropriate entities to handle projecting and tracking enrollment at study sites, monitoring the progress of clinical trials, and data management; whether existing entities could be engaged or adapted to carry out these functions for coordinated, large-scale emergency clinical trials.

COMMENT- "Because of the distributed ledger capabilities of blockchain and the ability to create public, closed, or hybrid models, our suggestion is to examine the combinations of models for the most optimal as requirements dictate."

2. b.,vi. **RFI-** Leveraging Community Based Networks

COMMENT- "Our work with matching FHIR elements to a blockchain and against the Uniform Data Standard format, especial UDS+, as a use of the proposed model is natural for FQHC's in research."

4. Emergency Master Agreement.

RFI- b. Additional terms for an Emergency Master Agreement that could be added or modified depending on the complexity of the protocol, and on other factors such as whether a private sector sponsor or an investigational agent is involved. It would be helpful to have input on terms such as the following:

i. Confidentiality.

COMMENT- "We are addressing this subset focusing on the field of electronic signature, confidentiality and it's assurance of validity. This also applies to providers or researchers and the prevention of fraud or hacking research data bases. Within the realm of licensing and credentialing, blockchain technology can be used to issue, share, and verify digital academic credentials. The Federation of State Medical Boards recently launched a pilot to test the use of Blockcerts, a blockchain-based open source standard, to create digital records for medical certification credentials. Regarding fraud protection, blockchain can publicly maintain a complete chain of custody."

RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Alex Cheng PhD - Research Assistant Professor of Biomedical Informatics, Vanderbilt University Medical Center

Paul Harris PhD - Professor of Biomedical Informatics, Biostatistics, and Biomedical Engineering, Vanderbilt University Medical Center

Introduction

We are writing to provide input to the Data for Emergency Clinical Trials RFI based on our work and experiences using the REDCap data management platform, developed and disseminated by our team at Vanderbilt University Medical Center. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) application programming interfaces (APIs) for data integration and interoperability with external sources.

We understand that a specific objective for this RFI is to gather information about whether there is value in a pilot or demonstration project to operationalize data capture in the near term, for example within 6-12 months of the close of comments on this RFI. We believe REDCap is an ideal platform to perform one or more interoperability pilots for emergency and pre-emergency clinical trials for the following reasons:

- REDCap is available at no cost to non-profit, government, and academic institutions which makes it accessible to groups reaching underserved populations.
- REDCap is well known in the biomedical research community. The platform is licensed and used at more than 6,000 academic, non-profit, and government organizations across 151 countries. REDCap supports many diverse research use cases, including clinical trials and public health monitoring.
- REDCap can support multi-center trials as a central collection platform.
- REDCap can also support multi-center trials through a distributed network of sites sourcing data derived from local EHR systems that can be asynchronously shared in identified or de-identified manner to a central data coordinating center.
- REDCap project-level metadata is easily transferred between REDCap installations, making data and project sharing via file downloads or via API interface with external systems straightforward.
- REDCap electronic case report forms (eCRFs) and patient reported outcome (PRO) instruments can be created independently or downloaded directly from an established Shared Data Instrument Library which includes thousands of downloadable validated instruments licensed without fee for immediate use by any research team using REDCap across the world (PMID = 23149159).
- REDCap is being used in the NIH-sponsored ACTIV-4 Host Tissue platform trial to test multiple investigational agents at many sites and also the ACTIV-6 remote trial platform. We recently conducted comparison experiments on the ACTIV-4

Host Tissue data showing that data coverage is high, data accuracy is improved, and coordinator burden can be reduced when automatically sourcing EHR data into eCRFs compared to traditional entry methods (JCTS Accepted for Publication - doi:10.1017/cts.2022.514).

- REDCap was used by the Oxford Vaccine Group to provide comprehensive support for AstraZeneca COVID vaccine trials.
- REDCap has an established collaboration with Epic which has resulted in a no-fee HL7 FHIR-based integration module using Smart on FHIR that is currently available in the Epic App Orchard. REDCap partner institutions have also hybridized the application to support FHIR-based EHR-to-REDCap connectivity in Cerner systems.
- REDCap's existing SMART on FHIR EHR workflow integration and FHIR-based data transfer methods are configured once by Health IT teams for use on as many projects as desired (PMID = 34298155). This design principle reduces the burden for institutions to install separate EHR integration products for each study or trial. At Vanderbilt.
- REDCap has the ability to rapidly map and enable EHR→ eCRF data transfer for prospective studies and trials and also harvest exist data for registry use cases. On a smaller scale, our team has also developed a pilot-phase external module that enables FHIR mapping for any data in a REDCap data collection instrument for data export purposes (e.g. REDCap as input for third-party interoperable data repository using FHIR payloads).

REDCap Alignment with RFI Workflow Scenarios

We provide detail below on how REDCap could be used in each of the tasks outlined in this RFI for multi-site clinical trials and ideas for potential pilots in domains relevant for this RFI. Notably, our responses and ideas are informed by real-world experience with our own data management projects and other researchers using REDCap at Vanderbilt and across the REDCap Consortium. The following sections outline how setting up a national multi-site emergency clinical trial with REDCap would be achievable:

1. A U.S.-level governing entity would oversee development of a clinical trial protocol for broad distribution across clinical trial networks and sites.

Government entity would commission a Clinical Coordinating Center (CCC) and Data Coordinating Center (DCC) to develop a protocol. Data collection instruments (including eCRFs and PROs) are built in REDCap based on protocol needs using the REDCap study authoring platform. When applicable, eCRFs and PROs can be exported directly from the REDCap Shared Library. Notably, REDCap eCRF authoring tools include real-time discovery and easily implemented common data elements (CDEs) from the NIH Common Data Element (CDE) Repository which houses all vetted NIH CDEs, including some USCDI data elements (https://cde.nlm.nih.gov/cde). Adding more USCDI data elements into the CDE Repository would result in automatic availability for CRF creation teams. Data elements can also be entered into REDCap directly for use before availability in the CDE Repository.

2. Study sites would enroll participants in the trial (potentially using software mechanisms that can alert sites to potential subjects for a specific protocol in a manner that increases the diversity of trial populations). Sites would obtain appropriate e-consents and authorizations from participants.

The DCC team could develop data collection centrally or in a distributed manner as described above. In our experience, local data sharing committees, privacy offices, cybersecurity teams, and health IT groups are comfortable connecting a local EHR system to a local EDC system, so our work to date has focused on a centrally managed process with local data sharing/curation prior to remote DCC data sharing. The DCC would build a harmonized project creation file created using specifications outlined by the government sponsor to create both a centralized REDCap repository and a local version of the study case report forms. Sites could use inclusion criteria triggers from the EHR to automatically add potential subjects to their local REDCap instance. Sites could use the REDCap-EHR FHIR interface to pull mapped study data directly from the EHR directly into REDCap. REDCap has a built in validated e-consent framework which sites could use to obtain consent and authorization from participants. Local site collection and management of eConsent forms can also be conducted using an established and validated REDCap eConsent module (PMID = 33244416).

3. Clinical trial data is typically sent to the trial sponsor through an electronic case report form (eCRF), which is the record of data that is required under the protocol to be captured for each trial participant. A data element in an eCRF is the smallest unit of observation for a particular subject.

Based on our experiences, it is unrealistic to expect complete coverage of eCRF data from EHR-derived data sources (doi:10.1017/cts.2022.514). That said, our REDCap EHR→ eCRF mapping tool is user friendly and enables rapid setup for automated transfer from EHR to eCRF when data are available. A decentralized model for site collection would also allow for local context specifications when use of standards is different across sites. Another important consideration in the transfer of data from the EHR to eCRF is whether adjudication is needed by a coordinator. We always recommend that initial transfers include an adjudication step to ensure expected results, but have built into REDCap workstreams the ability for coordinators to auto-adjudicate after validating the mapping/ and transfer process for a given study. For cases where eCRF data cannot be automatically pulled from the EHR, REDCap's eCRF instruments can be populated by a data entry staff member. In all cases (automated, semi-automated, manual), REDCap maintains secure audit trails necessary for project-level quality and attribution validation.

4. The eCRFs would be transmitted electronically via common APIs to the sponsor.

REDCap has an extensive set of API services available at the project level. In the scenarios described in this RFI, ,the sponsor or dedicated DCC would obtain API key from sites to allow for de-identified data aggregation with the sponsor. REDCap has native functionality which automatically strips identifiers and performs data shifting if desirable. The data model for each of the sites could be enforced by REDCap API interrogation of project metadata to ensure that harvesting of site data is streamlined and standardized.

5. The study site's health IT system would present the eCRF content to clinicians in a manner that expedites data collection and (ideally) fits within clinician workflows.

REDCap eCRF workflow is well established and has been used in thousands of studies conducted across the world. The learning curve is low and also helped by the fact that many clinicians and coordinators are already familiar with the platform based on experience with previous studies. The workflow of completing a CRF in REDCap is intuitive and familiar for clinicians. REDCap also has the ability to embed case report forms within Epic and Cerner EHR browser windows using SMART on FHIR methods developed and disseminated by our REDCap development team. We have seen in our work that some study teams opt for data entry directly in context of EHR utilization and others prefer to use REDCap as a sole front-end user interface with EHR data connectivity and exchange happening in the background and presented only when new data arrive that needs adjudication.

6. As the clinician obtains data elements to complete the eCRF, that data would be captured in the patient's electronic health record.

This expectation will take time to realize in a scalable manner. There are regulations and local policy constraints that prevent external applications from creating clinical records in the EHR. EHR software vendors have also been slow to build FHIR-based API endpoints allowing data ingestion. We have created workflow models to support automated creation/transfer/storage of REDCap-derived PDF document summaries within the EHR. We have also successfully worked with vendor-specific (Epic) non-FHIR APIs to selectively modify research study data within the EHR system (e.g. read or write participant status for an individual research study), but have not yet created a scalable and systematic method for setting up eCRF→EHR data transfer. As the field of real-world data for research progresses, we would like to tackle this problem.

7. The clinical trial data would also be sent to a central data repository or small set of data repositories for researchers to analyze. It would be sent via common APIs so that researchers can easily interpret the eCRF data elements. Commercial cloud solutions are likely to house the data repository or repositories. Nonetheless, we would like a solution that would work across multiple cloud vendors.

Existing REDCap APIs at each site can be used to pull data into central data repositories. These data can be de-identified by REDCap automatically prior to transfer. Additionally, if the central repository is a REDCap database itself, the sponsor can use an add-on external module called “API Sync” to automate pulling data from other REDCap instances. Our development team has worked with cloud engineering teams at AWS and Azure to create quick install/maintain versions of cloud deployment. Our operational teams have set up working installations in all 3 major cloud ecosystems and currently run an installation of REDCap in the GCP cloud environment that is part of a FISMA-Moderate program (https://allofus.nih.gov/) .

Response to Selected RFI items

1. United States Core Data for Interoperability (USCDI).

We seek input on how U.S. Government and external stakeholders might leverage USCDI and future extensions of USCDI standards (such as USCDI+, an extension that supports federal partner program-specific requirements) to support emergency clinical trial research. It would also be helpful to receive comment on areas in which additional extensions might be necessary.

We recommend that USCDI and USCDI+ data elements be made available on the NIH CDE Repository, even as new versions are being released. The NIH CDE Repository has APIs that allow for integration and dissemination to external systems like REDCap which would increase the access and use of the standards.

2. HL7 FHIR APIs.

We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-emergency phase, and in what areas further advances might be needed. Specific topics in this connection include:

a. Use of an API that supports FHIR Bulk Data Access to support clinical research; whether bulk data exports from EHR systems can be used to support certain clinical trial protocols.

There are currently limitations with how FHIR Bulk Data exports are created by EHR vendors using APIs. In Epic, patient lists must be created in the EHR before data are available to external systems through Bulk FHIR. The workflow we have created with REDCap queries data from the EHR FHIR APIs one patient at a time, but is automated and can run as a background process. This approach has been sufficient for the clinical trials we have run using the EHR FHIR integration, including those with hundreds of patients and thousands of data values. As Bulk-FHIR implementation evolves, we believe it will be possible to build additional modules and workflows that allow for faster and robust transfer, and we are interested in tackling this problem.

b. Use of the FHIR Questionnaire and QuestionnaireResponse resources to support clinical research.

Clinical researchers have typically used REDCap to collect questionnaire data. Attempts to push or pull data from the FHIR QuestionnaireResponse resource in Epic and Cerner are ongoing. In Epic, the workflow for feeding structured EDC data into the EHR via FHIR Questionnaire and Questionnaire responses (using recent release methods), but the dependencies and workflow are complex and project-specific. This is a good area for exploration using pilot studies which could then inform larger scalable solutions and we would be interested in tackling this problem.

3. SMART on FHIR APIs:

We seek input on how U.S. Government and external stakeholders might leverage SMART on FHIR APIs, and in what areas further extensions might be needed. It would be helpful to receive comments on:

a. The most promising ways to create SMART on FHIR technologies that are portable across different institutions and EHR systems, but also provide adequate functionality to support emergency clinical trial research.

The most direct way to collect clinical trial data from EHR systems is to create a custom SMART on FHIR application. However, this can be costly, time intensive, and difficult to maintain. In our experience, creating one-off applications for individual studies, especially where EHR data will be used, is laborious and can lead to significant delays as each institution must apply governance principles, prioritize work from Health IT, and clear other “authority to operate” hurdles. We would recommend, wherever possible, to use ‘build-once, use many’ integration designs like we’ve successfully designed, deployed and disseminated in REDCap.

b. Whether the portability of SMART on FHIR tools provides a way to reach institutions and sites that have limited information technology resources; any promising ways to use SMART on FHIR to expand clinical research into underserved settings

A SMART on FHIR app such as REDCap is less of a technical burden on hospitals' IT staff than other means of clinical data collection, such as dedicated teams building and maintaining specific research databases. We estimate that it takes about 20 hours divided between Health IT and SMART app developers to establish a REDCap connection with the EHR. Once connected, the integration can be used by many eCRF-driven or registry studies.

4. Clinical Decision Support (CDS) Hooks:

We seek comments on how the HL7 CDS Hooks specification might be used to support clinical research, for example by creating prompts within the practitioner workflow during interaction with patients; and any advances that might be needed to support the use case described above.

CDS Hooks have the potential to query data bi-directionally to/from SMART on FHIR apps so that data from external applications can be pulled into the EHR workflow. We have found that EHR vendors' adoption of CDS hooks lags behind their adoption of FHIR. As such, we have developed other ways to automate the exchange of data initiated by the EHR (as opposed to FHIR APIs where the data extraction is initiated by the external application). For instance, we have been able to use Epic's Event Driven Service Oriented Architecture (EDSOA) to trigger the transmission of medical record numbers from Epic to REDCap based on study inclusion criteria or study status. Once REDCap receives the MRN from Epic, REDCap can initiate automated extraction of other clinical data using FHIR. As CDS Hooks is maturing, we believe this would be an opportune time to build additional pilots which could then inform larger, robust scalable use and use cases.

5. Operationalizing protocols of varying complexity.

As noted above, emergency clinical trial designs could range from relatively simple protocols to more complex studies involving the evaluation of investigational agents. We would appreciate comments on the following topics:

a. Whether any of the tools described above might be particularly well suited for certain types of studies.

b. For example,

i. Whether a bulk FHIR API export could be used to gather data for a simple trial protocol that is relatively close to the standard of care for a particular condition.

We have conducted landscape assessment exercises assessing use of FHIR for streamlining the conduct of clinical and translational research (PMID = 34298155). We believe the potential for Bulk-FHIR methods is high, but currently immature in terms of implementation. Only very recent versions of Epic have the ability for bulk FHIR API exports. We have plans to evaluate bulk FHIR when it becomes available in VUMC's Epic instance in February 2023. In the meantime, we have been successful in pulling large amounts of data for many patients but one patient at a time in the background with REDCap. As mentioned above, this would seem an opportune space for pilot work which could inform larger scalable methods in support of data for emergency clinical trials.

ii. Whether a FHIR Questionnaire/QuestionnaireResponse or a SMART on FHIR form would be useful in capturing data for a more complex protocol, such as one that involves an investigational agent.

The Questionnaire and QuestionnaireResponse Resources in Epic have historically only able to export data captured in Epic questionnaires, which are not as flexible as other data capture tools designed for complex clinical trials. We typically use REDCap to collect questionnaire data and complement those questionnaires with data pulled from the EHR. Newer methods are now being deployed for bi-directional exchange using FHIR Questionnaire/Questionnaire Response resources, but dependencies are burdensome and not yet mature enough for building generalizable and scalable approaches. This would be a good area for pilot study exploration which could inform later scalable methods in support of data for emergency clinical trials.

c. Any technical limitations that we should be aware of regarding use of the above tools to operationalize clinical trial protocols.

We have demonstrated in an accepted manuscript (doi:10.1017/cts.2022.514) that automated data collection using the EHR FHIR API data extraction with REDCap provides high (but not complete) coverage for eCRF fields with better accuracy than data entered by study personnel. This demonstration was conducted on a complex, multi-site, multi-drug platform trial called ACTIV4 Host Tissue with real study participants and their study data. The case report form metadata as well as the mapping to EHR data elements can be exported from the study designer and shared with other study sites to upload into their REDCap instances. One limitation to operationalizing the standardized study structure at study sites is that not every EHR codes concepts (vital signs, laboratory results, diagnoses, etc) the same way. Therefore, even with the standardized case report forms, there needs to be some level of testing and remapping of EHR data at every study site. Our development team has created a solution where local-context mapping can be performed and applied to the standardized metadata used for describing EDC and PRO data elements.

6. Consent, de identification, return of results.

The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution.

a. In light of this, we seek comment on how the tools described above can be used to obtain, collect and/or manage any required informed consents and/or authorizations from patients or individuals in accordance with applicable regulations.

REDCap has a built-in e-consent framework (PMID - 33244416) where participants can sign an online form to indicate consent. The e-consent framework has been validated and used widely in both remote and in-person clinical research studies as an e-consent form can be sent to a participant email address or can be filled out in-person with a member of the research team. Data collection and transmission to the central repository can be set up to be contingent on a completed e-consent form.

b. We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information. It would be helpful to receive comments on which deidentification and protection approaches are sufficiently mature to support a pilot effort in the near term.

We have experience working in scenarios where data needs to be de-identified before being sent to the DCC. In these cases, we mark all safe harbor variables as identifiers in REDCap so that they are automatically excluded from data export. REDCap has automated methods which support automated longitudinal sharing using a derived unique participant study identifier and date shifting by a person-specific derived random number (0-364) prior to export to meet most de-identification transfer requirements. Removing identifier variables and date shifting in REDCap is possible both through API data transfers and through flat file downloads. These methods are mature in REDCap, but we have not yet built methods to deal with de-identification of unstructured (e.g. notes) data. This would be an interesting area for pilot development and would also inform development of new scalable methods like automated concept extraction from clinical notes in support of emergency clinical trials.

c. Ideally, patient authorization would allow clinical trial data to be used for additional research beyond the original study. We would appreciate input on how the content collected for consent and authorization as well as the interfaces with deidentification technologies should be designed to enable flexible and responsible reuse of clinical trial data.

Several clinical research organizations have released templates for informed consent form language for secondary use and data repository research are available.

d. We seek comment on any technical capabilities that could support return of results to study sites or participants, where appropriate.

Previous studies have used REDCap to deliver return of results to participants (PMID = 30239733). These can be done as an aggregate view (i.e. graphs and tables summarizing results from the study across all participants) through a dashboard with a public or private link. Using integrations with SMS + e-mail services, REDCap can also send participant specific results or links to results directly to participants as long as privacy considerations are met. For longitudinal studies employing frequent engagement with participants, our MyCap mobile application can collect data as well as deliver secure in-app messaging (PMID=35673353).

e. We seek comment on any regulatory or ethical guidelines that are relevant to patients' consents and authorizations under the use case described in this RFI, and on ways in which technical solutions might help ensure adherence to applicable regulatory or ethical guidelines.

Any study documents will have to be approved by each site's IRB or a single IRB. It is likely that individual sites will request site-specific modifications to the informed consent forms and process. Using a single IRB management system such as IREx (a platform also developed by Vanderbilt and funded by the NIH) can help to ensure that proper regulatory and ethical procedures are followed. IREx is highly utilized and well trusted by clinical researchers and has been used in 482 studies. https://www.irbexchange.org/p/ (PMID = 35574155).

7. User interface and experience. With all of the above technologies, we seek input on:

a. The best way to optimize the experience of health care providers, administrators, and other users, so as to maximize the utility and uptake of the product.

One of the primary benefits of using REDCap for these rapid emergency trials is that nearly all academic medical centers are already using the system and should be familiar with its interface and capabilities. As such, study designers can quickly disseminate standard case report form metadata and start data collection as soon as possible.

b. To the extent a particular form, app or other tool requires input from a health care provider or other user, the best ways to increase the likelihood that users will actually provide that input. It would be helpful to receive comments on methods that are available for completing empty fields after the fact, or otherwise managing any missing data.

REDCap users can tag other users to check inaccurate data or complete missing data through a feature called the “Data Resolution Workflow”. Study personnel and health care providers are much more likely to respond to completing missing data when they are assigned the responsibility by a sponsor or supervisor. REDCap also has an embedded customizable data quality module that reports on frequent cause (missingness, branching logic errors) and programmable complex scenarios (e.g. visit date ordering, multi-dimensional data checks like pregnancy in males and/or outside expected age limits). Regular review and resolution of quality issues during pilot phases of a project can ensure full-team discussions to help solidify SOPS before larger trial launch.

c. For clinicians and health IT users: what existing tools, apps, or processes you have found most usable and why.

REDCap is an “existing” tool for over 2.3M users at more than 6,000 organizations in 151 countries. Part of what makes it popular, aside from it being available at no cost to non profit, government, and academic institutions, is that it designed specifically for clinical research. It is also highly flexible and gives researchers control over their own eCRFs, PROs and study design with an easy-to-use interface.

8. Capturing data elements required for clinical trial protocols.

a. We seek comment on the most promising technical approaches that would leverage common APIs to translate a particular clinical trial's data elements into data elements captured by user-facing tools (e.g. FHIR Questionnaire feeding into a SMART on FHIR form or application).

Since researchers are familiar with REDCap’s eCRF and PRO questionnaire authoring tools and clinical trial management tools, we recommend that trials create the questionnaires in REDCap. Through a REDCap module, questionnaire (and other clinical data) can be mapped to FHIR resources and exported in a FHIR bundle for ingestion to other servers such as a central repository. Development of this external module (called FHIR Services) is ongoing and could be adapted for this use case. The “FHIR Services” module is different from “Clinical Data Interoperability Services”, which is REDCap’s standard method for supporting FHIR integration with EHR systems. The Clinical Data Integration Services module is already mature software and in use at many institutions across the country. (PMID = 34298155).

b. If a tool such as a FHIR Questionnaire, FHIR QuestionnaireResponse, or SMART form or app is used to capture required data elements in this way, we seek comment on whether that creates an effective method for “pushing out” a research protocol to investigators and sites.

The current specification for the Questionnaire resource in FHIR cannot adequately describe a trial protocol. Details such as user roles, the frequency and timing of events, and the method of sending the questionnaire are difficult to define in FHIR but easy to set up in REDCap. REDCap projects can be configured so researchers are guided to “do the right thing” when setting up data management services in support of a new

protocol. Moreover, these project configurations can be exported as an .xml file and imported into any sites' REDCap instance so that each site has a matching project.

c. It would be helpful to receive comments on how best to ensure compliance with regulatory requirements for eCRFs when designing interfaces for data capture.

The REDCap Consortium has a standing volunteer group sharing knowledge and methods for local validation. Several institutions (including Vanderbilt) have shared documentation related to local validation to fulfill requirements for 21 CFR Part 11 compliance. That said, the local validation process is onerous and not conducive to rapid deployment of researcher-informed features. We are in the process of reengineering the process using a more technical framework for creation and automated deployment and documentation of validation scripts. The development and evolution of this "rapid validation" framework might serve nicely as a pilot project in the Data for Emergency Clinical Trials space, ultimately resulting in new ways to support platform validation for studies using REDCap and also as a generalizable/innovative methodology for potential use by other technology teams.

9. TEFCA and QHINs and 10. Emerging technologies. See other sections.

11. Pilot or demonstration project.

We seek comment on how the U.S. Government can best work with external stakeholders and developers to develop a pilot or demonstration project that will operationalize clinical trial data capture and serve as a basis and model for data collection in the event of an emergency. This pilot or demonstration project could also potentially support clinical research in the pre-emergency phase. Specific topics include:

a. Whether data can be managed through a central repository or small set of central data repositories; options for cloud-based data storage.

We believe that REDCap would be an ideal platform to conduct a pilot for emergency clinical research studies and have listed numerous ideas for pilots in preceding sections of this report. For pilot work specific to central/site data collection and transfer, we would envision setting up a REDCap instance in the cloud (AWS, GCP, or Azure) as a central repository. Existing REDCap tools could be used to automate data transfer from contributing sites' REDCap instance to the central repository instance used to harmonize all data.

b. Technical options that might hold promise in the short term to enable researchers from diverse locations to analyze the data collected from multiple clinical trial sites. We also seek comment on any additional options that should be considered in the long term.

REDCap is used at thousands of institutions across the US, large and small, urban and rural. It would be perfect for a short term pilot but could also be considered as a long term solution. REDCap does not have native tools for complex analysis or multi-project dashboards, but does have an extensive set of APIs and traditional user interface (point and click) modules in place for export of data to common statistical platforms (e.g. SAS, Stata, SPSS, R), data science tools (e.g. Python), and rapid sharing with business intelligence tools (e.g. Tableau).

12. Specific commercial capabilities.

REDCap Consortium and no-cost licensing terms for adopting sites are listed on the REDCap (https://projectredcap.org/).

Data Collection for Clinical Trials RFI

Comments and Response

Addressed to: *datacollectionforclinicaltrials@ostp.eop.gov*
Submitted by: Samir Jain, Sr Director, EHR Solutions, Medidata Solutions (samir.jain@3ds.com)

About Medidata Solutions

Medidata Solutions, a Dassault Systèmes Company, is the leading provider of clinical trial management, electronic data capture, and decentralized clinical trial solutions. With over 1700+ customers and partners, and over one million registered users, Medidata is the most trusted platform for clinical development, commercial and real world data.

The author offers this response from the perspective of a Health IT and App Developer.

RFI Response

1. United States Core Data for Interoperability (USCDI). We seek input on how U.S. Government and external stakeholders might leverage USCDI and future extensions of USCDI standards (such as USCDI+, an extension that supports federal partner program-specific requirements) to support emergency clinical trial research. It would also be helpful to receive comment on areas in which additional extensions might be necessary.

As developers of Electronic Data Capture (EDC) and Clinical Trial Management Software (CTMS) solutions, we continually probe our customer base and stakeholder groups for new, novel, and more efficient ways to complete workflows typically associated with data capture during a clinical trial. Overwhelmingly, our research site stakeholders have expressed frustration with today's highly manual and time consuming source to EDC transcription processes. Many solutions leverage manual mapping efforts that must be repeated for each study protocol / site EHR integration combination. These efforts become burdensome, prohibitively expensive, and infeasible at scale.

The goals of clinical data capture, when done outside of the confines of a research study, are different from the goals of data capture for research. Clinical data are meant to be captured with as much precision and uniqueness to the patient's specific condition, activities and plan as possible. These then help drive continuity of care, patient history, and billing workflows. Research data, alternatively, are meant to capture the experience of an individual as part of a larger cohort, as experienced through the lens of a well defined study protocol. There are inherent differences in the need for resolution and fidelity of data, given these use cases. We feel it may be infeasible to change USCDI design without overly burdening EHR developers to support data collection and codification standards meant to assist in research (IE, more closely aligned to CDISC specifications).

However there is an opportunity to align and homogenize EHR developers' interpretation of USCDI and specifically, their implementation of FHIR specifications. We believe that ONC's existing Final Rule timeline accounts for inherent differences in implementation by including real world testing phases. The goal that we advocate for, is the ability to confidently expect homogeneity between EHR implementations of FHIR, including code systems used, to ultimately create efficiencies and turn-key reusability across FHIR to CDISC and eCRF mappings.

2. HL7 FHIR APIs. We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-emergency phase, and in what areas further advances might be needed.

FHIR and the alignment around a single version and interpretation, presents a large opportunity to quickly integrate data across multiple disparate systems, for purposes of broad-scale research. Specifically:

a) The use of Bulk Data Access to support study participant identification and pre-screening for inclusion / exclusion criteria
b) Access to captured clinical data for purposes of completing eCRFs across a variety of organizations and EHR systems
c) In-workflow integration of data capture tools

Key barriers to date have been:

a) Lack of a single, trustable directory of FHIR endpoints across organizations
b) Lack of a common agreement for access (the need for individual data usage agreements for each integration)
c) Variations in FHIR implementations

3. SMART on FHIR APIs: We seek input on how U.S. Government and external stakeholders might leverage SMART on FHIR APIs, and in what areas further extensions might be needed.

As mentioned above, a key opportunity enabled by FHIR is the ability to integrate data capture tools into existing workflows, via SMART on FHIR. The two major advantages of SMART on FHIR are a) it's relative simplicity and ease of use and b) it's wide-scale adoption and enablement by EHR vendors, and inclusion of SMART on FHIR as the framework for developing apps within their ecosystems.

When developing a SMART on FHIR app, we believe the keys to success are:

1) Leverage "pre-fillable" fields wherever possible to assist the user
2) Appropriate and sensical launch points within the EHR workflow, along with user experience continuity when launching a SMART on FHIR app

In addition, SMART on FHIR, as a standard, does not limit itself to EHR implementation. Specifically, there is no reason why CTMS systems cannot also implement SMART on FHIR launch capabilities. As such, smaller, independent research sites that operate without a traditional EHR may gain access to functionality delivered via SMART on FHIR, given their CTMS vendor's willingness to implement the standard.

Opportunities for ONC include sponsorship and maintenance of a:

1. Single, easy to follow, up to date SMART on FHIR quick start guide (for both server and client models). This tutorial should assume limited working knowledge of FHIR.
2. Github repository for Server and Client app examples
3. Easy online testing toolset to test both Server and Client apps

4. Clinical Decision Support (CDS) Hooks: We seek comments on how the HL7 CDS Hooks specification might be used to support clinical research, for example by creating prompts within the practitioner workflow during interaction with patients; and any advances that might be needed to support the use case described above.

CDS Hooks is a perfect candidate to embed both candidate identification/recruitment, as well as eCRF workflows. However, wide-scale adoption of CDS Hooks by EHR Vendors remains elusive. While some major EHR vendors (such as Epic Systems) have enabled CDS Hooks capabilities, and configurability by their customers, many others have not yet made CDS Hooks and configuration frameworks widely available to their customers.

There is an opportunity for ONC to include the implementation of CDS Hooks as a CEHRT criteria, which will de facto drive wide scale availability.

5. Operationalizing protocols of varying complexity. As noted above, emergency clinical trial designs could range from relatively simple protocols to more complex studies involving the evaluation of investigational agents.

The success of any integration technology is driven by its wide scale applicability and adoption. To date, FHIR Questionnaires have neither been mandated, or commercially driven to be implemented with EHRs. Without specific regulation, there is a low likelihood of broad and consistent support of FHIR questionnaires within EHRs. As such, there is a much higher likelihood of success in implementing an embedded workflow via SMART on FHIR.

In this case, the EDC (or other embedded app) vendor does not gain much by implementing the eCRF in FHIR Questionnaire / Response format. Success, rather, hinges on the vendors ability to embed itself in clinical workflows effectively (via SMART on FHIR), drive an efficient user experience to encourage form fills, and transform input into a data model that can then be combined and leveraged at scale.

6. Consent, deidentification, return of results. The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution.

Medidata offers the Medidata Link product to help customers tokenize patients in partnership with partners such as Datavant and Health Verity, and leverage real-world datasets aligned to these tokens. As ONC considers a potential approach for a pilot effort, they should look at existing adoption, scale, and ease of integration with other participants in the pilot.

7. User interface and experience.

In order to maximize engagement with clinical users, the following should be considered:

1. Embeddability within existing tools - we have found that being minimally intrusive to existing workflows increases adoption and engagement. For example, “swivel chair” and dual application workflows should be avoided. In unavoidable, at minimum, cross application SSO is required to prevent funnel drop offs.
2. Prefill of information - whenever possible, information such as subject ID, visit dates, author / site information, and clinical information should be pre-filled, or easily selectable by the user.

Many EDC systems already contain workflows for form field validations, error and anomaly detection, and data verification. Any pilot program should aim to take advantage of these existing capabilities, rather than recreating these workflows specifically for this project. As such, embedding existing EDC experiences within EHR workflows via SMART on FHIR, while accounting for the user experience considerations highlighted above, would be an optimal approach.

8. *Capturing data elements required for clinical trial protocols.*

The points highlighted in this question have been discussed above.

9. TEFCA and QHINs. As noted above, TEFCA is in the implementation phase at this time. In the future, the TEFCA QHINs are expected to support implementation of the FHIR APIs (see the ONC Recognized Coordinating Entity's January 2022 FHIR Roadmap for TEFCA Exchange [6]). We would appreciate comment on the opportunities and challenges regarding development of API implementations toward the use case described above, particularly given the current status of TEFCA and QHIN participation.

TEFCA, and specifically the Common Agreement present an exciting opportunity for Clinical Research, if made available. Key barriers to integrating with and using EHR data have been the need for:

1. Individual IT projects at each site, with development, testing, and project management
2. Site IT leadership relationships
3. Individual site data usage and sharing agreements

Moving towards a “mesh network” approach of QHIN - QHIN connectivity, coupled with a common data sharing agreement significantly reduces these barriers, so long as there is a participant willingness (or mandate) to respond to Clinical Research purposes. The creation of this purpose of use will be a key to Medidata’s investment in a research focused integration strategy that centers around TEFCA.

One approach that may improve short-term feasibility would be to separate out use cases where new and novel information is discovered (such as candidate identification, post-marketing surveillance, and longitudinal follow-up), from others where users already have access to information and are hoping to create operational efficiences (such as EHR to EDC where a user is manually transcribing data). The latter may be

easier to adopt, while the former may take more time and consideration of privacy and consent.

While this response does not provide any legal review or opinion, The guidelines for Uses and disclosures for public health activities (45 CFR § 164.512 (b) (1) (i) and 45 CFR § 164.512 (b) (1) (iii)) may be applicable. QHINs may benefit from more specific guidance as to when these regulations are applicable to clinical trials and sponsors, allowing them to more effectively self-regulate their participants. Failure to do so may result in QHINs taking a more risk-averse regulatory interpretation and overly restrictive approach.

10. Emerging technologies. We welcome comments on any future technological developments we should anticipate. Relevant technical developments include but are not limited to differential privacy; federated machine learning; other technologies referenced in the recent OSTP RFI related to privacy-enhancing technologies (PET) (*see***Federal Register:** Request for Information on Advancing Privacy-Enhancing Technologies); and technologies outside of the PET space.

No response provided.

11. Pilot or demonstration project. We seek comment on how the U.S. Government can best work with external stakeholders and developers to develop a pilot or demonstration project that will operationalize clinical trial data capture and serve as a basis and model for data collection in the event of an emergency. This pilot or demonstration project could also potentially support clinical research in the pre-emergency phase.

Medidata would be interested in supporting ONC in developing a pilot / demonstration project to demonstrate the feasibility of an emergency clinical trial. Our recommendation would be to leverage capabilities that are easily configured within an EHR, such as SMART on FHIR, and involve a diverse set of participants that represent EHR vendors, both independent and affiliated research sites, and EDC vendors.

Commercial, public-private partnerships, or ONC provided funding would all be viable pathways. Participants will likely want to understand the specific decisions / assumptions being tested, and the likely outcome and result of a successful demonstration.

12. Specific commercial capabilities. Commenters who are developing a technology or product that might be relevant to any of the topics set forth above are welcome to

include a description of that product. Comments about a specific technology or product should be limited to three pages or less.

Medidata recently announced its Rave Companion product, which facilitates clinical data transfer between various source systems into Rave EDC. Rave Companion works in two different modes: Integrated, where data is pulled directly from an EHR system through a variety of interfaces and integration mechanisms, and Standalone, that allows the user to pick and choose which data elements from various systems to bring in. More information about Rave Companion, including a video demonstration, can be found at: https://www.medidata.com/en/clinical-trial-products/clinical-data-management/rave-companion/

Input from AccendoWave - a health information technology (health IT) company.

Overview

AccendoWave is a Pain Data Company with machine learning pain measurement & management technology. A Health Equity Solution and Top 50 Remote Monitoring Company, AccendoWave benchmarks pain data (specialty, gender, age) to eliminate bias, improve outcomes and reduce health care costs.

Strategic Relationships

Incubated by Hospital Corporation of America (HCA) - Samsung and AT&T are AccendoWave's partners.

Value Proposition

AccendoWave correlates your brain waves with your perception of pain to:

1. Validate your pain so your pain is believed and your care and outcomes improve.

2. Provide objective pain data to reduce the cost of care.

EEG Research

1. https://pubmed.ncbi.nlm.nih.gov/32591813/

2. EEG Research Study Attached

Clinical Trial Gap & Data Partnering

Today in our clinical trials we have a significant gap. No objective pain technology and objective pain data is being used - so there is no ability to ensure consistency in pain measurement and pain data from one clinical trial site to another.

AccendoWave has begun to partner/license objective pain databases. Objective Pain Databases can now be found on the Datavant platform. Here a link to Datavant: https://datavant.com/
AccendoWave has nine benchmarked objective databases on pain (Maternal Health, MSK, Oncology, Women, Seniors, Adults, etc). If desired, objective pain databases can be customized for each specialized clinical trial population.

Value Proposition

Pain is the primary reason patients access health care - so pain is a primary driver of health care cost. Pain levels don't show up in claims data nor can objective pain levels be extracted from EHR's. So Employers/Health Plans/Hospitals/Government have no visibility into or data insights on the primary driver of health care costs.

Health care has a pain bias problem. Pain is one of the most subjective measures in all of medicine - and thus subject to bias. There is an extensive body of research that shows bias: women and pain, ethnicity and pain, and seniors and pain. For many in these groups - their #1 complaint is that their pain isn't believed.
This is manifesting as less pain medicine after C-sections and later stage cancer diagnosis.

Recognition

UCSF 2021 Digital Health Awards – Quarterfinalist New Health Application of AI
UCSF 2022 Digital Health Awards – Quarterfinalist Best in Class – Remote Diagnostic Tool or Device
AVIA Health Top 50 Remote Monitoring Company (Innovation Platform Used by 50+ Health Care Systems)
Milken Institute – Partnering for Patients: https://milkeninstitute.org/video/employers-mental-health-addiction
PCORI Board of Governors 2022 Nomination (Martha Lawrence, CEO AccendoWave)

Data Collection for Emergency Clinical Trials and Interoperability Pilot Response to the White House Office of Science and Technology Policy

Oracle America, Inc.

Document Number: 2022-23489 | January 27, 2023 | 5:00 PM (ET)

Submitted to:
White House Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, D.C. 20504
Attention: Grail Sipes, Assistant Director

Submitted by:
Oracle America, Inc.
500 Oracle Parkway
Redwood Shores, CA 94065
Jerrold Johnson, Applications Sales Representative

Oracle America, Inc.
500 Oracle Parkway
Redwood Shores, CA 94065

January 27, 2023

Grail Sipes, Assistant Director
White House Office of Science and Technology Policy
Executive Office of the President
Eisenhower Executive Office Building
1650 Pennsylvania Avenue
Washington, D.C. 20504

Dear Ms. Sipes:

On behalf of Oracle America, Inc. (Oracle), thank you for the opportunity to respond to the White House Office of Science and Technology Policy (OSTP). We are proposing Oracle Health and Life Sciences Data Collection for Clinical Trials to address OSTP's desire to establish a US-level governance structure that can participate in emergency research, both domestically and internationally, as well as coordinating large-scale clinical trials that can be efficiently carried out across various institutions and sites to address disease outbreaks and other emergencies.

Our response explains how Oracle can help. Indeed, since 1977, we have helped hundreds of thousands of customers of all sizes around the globe simplify their processes by engineering hardware and software to work together. We drive transformation inside your industry with dedicated vertical organizations with deep domain industry expertise to provide best-of-breed technologies to help solve the most complex business problems.

Oracle offers a complete technology stack in the cloud, on premise, and in the data center. Our stack of products gives customers complete deployment flexibility and the unmatched benefits of application integration, powerful performance, high availability, scalability, advanced security, energy efficiency, and low total cost of ownership. We help develop strategic, efficient processes by adopting technologies that enable health sciences companies to provide reliable, secure, and scalable technologies and processes that deliver results for their customers.

In fact, Oracle's cloud products help businesses, health sciences companies, and public institutions modernize, innovate, and compete in today's digital world. With this modern cloud, OSTP can meet your organization's objectives more quickly and efficiently.

In addition, we not only provide robust products, but Oracle also works with you on every step of the digital journey. OSTP will benefit from Oracle's customer support services and can also take advantage of optional services, such as consulting, training, upgrade support, and financing. We will help you get the most out of your Oracle products so that you can meet your business objectives.

We can also help implement your Data Collection for Emergency Clinical Trials and Interoperability Pilot. For implementation, we are proposing Oracle's Health Sciences Consulting (HSC) Services. Our experts know the Oracle products, your industry, and

the project pitfalls to avoid. In addition to implementation, Oracle HSC helps customers with architecture, planning, upgrade, migration, and expert services across the Oracle stack.

We value our growing relationship with OSTP and are excited to enhance it. Please feel free to contact me if you have any questions or would like further information. I can be reached at +1.443.756.8641 or via email at jerrold.johnson@oracle.com.

Sincerely,

DocuSigned by:
Jerrold Johnson
5D08FE03C9DF41B...

Jerrold Johnson
Applications Sales Representative

Table of Contents

Response Guidelines

Corporate Entity

This response is being made by Oracle America, Inc., a wholly owned subsidiary of Oracle Corporation. All responses reflect information concerning Oracle Corporation (hereinafter referred to as Oracle) except where otherwise indicated as being information of Oracle America, Inc. (hereinafter Oracle).

Understanding Oracle Terminology

Oracle understands the task ahead for OSTP to review and compare responses for your project. We believe both you and Oracle benefit from a common understanding of terminology. We have included “Appendix B: Definitions and Abbreviations Used in Oracle’s Proposal.” Please refer to this appendix for further details about what you can expect from Oracle should we win your business.

Response Validity

This response shall remain valid until February 28, 2023, unless otherwise mutually agreed, in writing, by Oracle and OSTP.

1. United States Core Data for Interoperability (USCDI)

USCDI can be leveraged to identify relevant data that typically support emergency clinical trials for enhanced preparedness. However, it would not necessarily provide a fast path to addressing additional data critical to an emergency clinical trial where such data is not yet part of USCDI. USCDI may not fully support Health Level Seven (HL7) Fast Healthcare Interoperability Resources (FHIR), and HL7 FHIR US Core in particular. For example, USCDI V3, published in July 2023, will not have the necessary HL7 FHIR US Core support until Q2 2023. We anticipate that this lag will persist. Thus, we recommend that the primary focus should be on ensuring the relevant data is in FHIR US Core as soon as it is known, and that as part of an emergency preparedness process for emergency clinical trial, there is a defined process to rapidly update/augment HL7 FHIR US Core when needed so that it can be deployed to the relevant data sources. HL7 FHIR US Core is emerging as the de facto data set that is HIT certified under the Office of the National Coordinator for Health Information Technology (ONC)'s certification program and must support to the extent that the data source manages data that is in scope of HL7 FHIR US Core.

We suggest that, rather than have a USCDI+ extension for emergency clinical trials, this be addressed through USCDI to help ensure that the HIT that manages such data would be able to rapidly expose the data within the defined access authorities.

2. HL7 FHIR APIs

HL7 FHIR based application programming interfaces (APIs) are well suited to address ad-hoc, emergency queries whether for incremental public health surveillance data needs, research, or emergency data access. HL7 FHIR US Core is the primary data set that certified HIT needs to support. Thus, a reasonable scope to expect is the largest possible set of relevant data sources supporting the certified HIT-managed data. In combination with Bulk Data export, which is also a capability that certified HIT must support for the data they manage, critical tools will be in place to support very targeted, individual data access to augment data already available as well as data sets for larger patient cohorts.

The HL7 FHIR questionnaire and questionnaire response resources are appropriate mechanisms to query the data of interest as well. We strongly recommend using these capabilities that are already in progress within HL7 Da Vinci to support data for prior authorizations. Additionally, we recommend considering the approach used by the electronic Case Reporting Now where, in both cases, the Cassandra Query Language (CQL) is used to define the data of interest and the requesting method. Alignment with these approaches will greatly enhance the ability to rapidly expand on additional data needs particularly if they are already covered by HL7 FHIR US Core.

We must note that APIs provide access to real-world data. This means, to the extent that emergency clinical trials rely on highly curated data, particularly additional data otherwise collected, the available APIs would not necessarily have access to the additional data specifically being collected for the trial. Such data could be collected using FHIR-based apps using SMART app capabilities enabling collection directly from the user as described in the next section.

3. SMART on FHIR APIs

The use of HL7 FHIR-based applications enable an intermediary approach between the requester and the data source. The application can facilitate the necessary data gathering using existing HL7 FHIR US Core-based APIs and Bulk Data access capabilities while using SMART to enable interaction with a user directly for any data that cannot be automatically retrieved (whether by using HL7 FHIR US Core-based APIs or proprietary access paths). The Da Vinci ePrior Authorization approach, as well as the HL7 Situational Awareness for Novel Epidemic Response (SANER) and Data Exchange for Quality Measure (DEQM) approach enable such HL7 FHIR-based applications. We suggest that these approaches be considered for an HL7 FHIR emergency clinical trial implementation guide that provides the flexibility for a source to directly provide the relevant data, including interaction with users (whether clinicians or patients), or a FHIR-based application can facilitate the data collection.

4. Clinical Decision Support (CDS) Hooks

CDS Hooks are intended to facilitate triggering of actions/decisions within a relevant workflow. For example, one can use CDS Hooks in an order or referral workflow to determine and then initiate an interaction with a payer to obtain the necessary prior authorization supporting information. Within the context of an emergency clinical trial, such triggers could be considered for patients participating in that trial as they are registered, relevant orders are placed, results are received, diagnoses are documented, etc. Another example is originating an activity to fill out the appropriate case report forms (CRFs) in the emergency clinical trial. Thus, CDS Hooks should be considered as a tool to identify opportunities where data can be shared proactively by the relevant data source with the target research system. Sharing the patient cohort, including updates and essential triggers will be critical to enable a quick response to new emergency clinical trials.

5. Operationalizing Protocols of Varying Complexity

To support varied protocol needs, systems and standards must be robust and flexible. Moreover, rapid setup is critical. Oracle Clinical One is a mature and proven solution ready for all types of studies. Visits and events are configurable to accommodate study designs from simple protocols to the most complex visit schedules with branching and cycling. Similarly, when creating forms, we offer a variety of form design types, data collection item types, and configurable dynamics that will meet the need of your protocol. Within Study Builder, studies can be implemented with complex source data verification and risk-based monitoring strategies. Clinical One was built with the intent to support adaptive trial designs out-of-the box. Mid-study changes are a major differentiator. There is no migration of data, no impact to end-users, no downtime; changes can be built, tested, and implemented rapidly. Furthermore, Clinical One’s open API architecture allows for an easier integration with third-party vendors, thus simplifying data ingestion for any type of study through the Digital Gateway. In addition, Clinical One study definition includes a user interface to define the randomization scheme. There is no limit to the

number of stratification factors. The system supports dynamic and fixed block assignment, minimization randomization, and multiple randomizations in a single study.

Oracle Data Management Workbench (DMW) provides the broad clinical data management capabilities to manage the variety of data sources emergency clinical trials will need, including the above-mentioned Clinical One site-entered data. DMW can be configured quickly using a library of pre-built study information, so researchers can aggregate, transform, review, and clean data from numerous sources in a centralized location.

We note that, in addition to the FHIR tools indicated, FHIR Bulk Data should be considered because support for FHIR Bulk Data is starting to emerge as well with the rollout of HIT certified to the 21st Century Cures Act related certification criteria. FHIR Bulk Data, the use of questionnaire, questionnaire response, and CQL provide the necessary tools to support the clinical trial data needs from simple to complex, by individual patient or in large volume, and collect them accordingly. The main consideration in choosing the appropriate approach is the amount of data, as well as the extent to which the data is readily available or needs to be captured specifically for the emergency clinical trial at hand. Patterns related to using these capabilities are emerging in HL7 FHIR Accelerator efforts such as Da Vinci, HELIOS, and Vulcan, which should also be considered. Alignment enables easier and quicker deployment of new emergency trials as they can build on infrastructure and capabilities already in place. Developing essential implementation guides will be critical to ensure the various FHIR capabilities are used consistently across the data collection workflows and the myriad of contributing HIT that are at varying degrees of FHIR adoption.

6. Consent, Deidentification, Return of Results

a. Tools to Obtain, Collect, and Manage Informed Consent

At Oracle, we believe digital and decentralized clinical trials are the new frontier of clinical research and provide the best opportunity to bring lifesaving therapies to patients faster and cheaper. Electronic consent documentation is typically collected and managed at the point of enrollment and kept in auditable form by accountable organizations engaged in clinical research. Centralized data repositories containing patient data may offer audit of consent status and may contain original consent documents, but do not typically allow research participants to dynamically interact with their prior consent (to either allow additional data sharing or revoke prior consent).

In a central data repository, like the one described above, the ability for participants to dynamically interact with their consent status is recommended as a method to increase trust and encourage participation from the broadest population. The National Institutes of Health (NIH) “All of Us” program has developed a mobile application that is a meaningful precedent for an interactive consent system.

b. Managing Protected Health Information

Section 164.514(a) of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule provides the standard for de-identification of protected health information. Under this standard, health information is not individually identifiable if it does not identify an individual and if the covered entity has no reasonable basis to believe it can be used to identify an individual.

Sections 164.514(b) and (c) contain implementation specifications that a covered entity must follow to meet the de-identification standard.

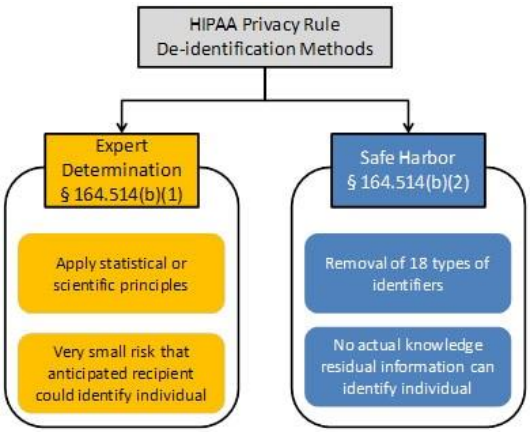


FIGURE 1: METHODS TO ACHIEVE DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION

Expert Determination Method – A person with appropriate knowledge and experience with generally accepted statistical and scientific principles and methods for rendering information NOT individually identifiable applies these principles, documents their method, and signs off that the risk of re-identification is small.

Safe Harbor Method – Automated or manual removal of 18 specified types of identifiers (e.g., names, geography, telephone numbers, email addresses, social security numbers (SSNs), biometric identifiers, account numbers, etc.).

Both de-identification methods are currently implemented in Oracle Cerner's de-identified datasets including our real-world data assets. In addition to adhering to de-identification standards, we recommend the use of an honest broker partner such as Datavant to tokenize de-identified datasets to facilitate clinical research and other secondary use cases of data while maintaining the privacy of clinical research participants. Tokenization enables enhancement of data sets and recombination of clinical research data while maintaining participant privacy.

c. Data Collection Design

The HIPAA privacy rule (specifically, 45 CFR 164.501, 164.508, 164.512(i)) establishes the conditions under which protected health information may be used or disclosed by covered entities for research purposes. A covered entity may always use or disclose health information for research purposes which has been de-identified (in accordance with 45 CFR 164.502(d), and 164.514(a)-(c) of the rule). The entity must specify the implementation specifications for valid authorizations for a covered entity to allow the release of protected health information for secondary use cases including research beyond the original study. Flexible re-use of clinical trial data requires patient authorization for secondary data-use:

- Explicit patient authorization for continued secondary data use (expiration at a set date far in the future or at study close)

- Transparency on the use case, who will access, and how data will be shared
- Respect participant preferences – allow customization of opt-in/opt-out based on individual comfort. Electronic means to opt-out at any point in the future.

d. Technical Capabilities

In general, returning results to study sites is well supported through cloud-enabled software that is used to manage and collect data for research studies. This functionality is often facilitated by user-based role assignment so that site level users can access data and functionality that is pertinent to their use case but are limited from data and functionality that extends beyond the scope of an individual's role in the project. It is becoming increasingly common practice for sponsored clinical research from for-profit and non-profit entities return reports to participants that are digestible by a lay audience and convey meaning and context to program participants. This type of engagement has shown to improve participant retention and benefit outreach, engagement, and recruitment activities.

e. Regulatory and Ethical Guidelines

Expeditious conduct of emergency trials while maintaining compliance with regulatory directives and ethical guidelines is best supported by technology enabled decision trees based on 'study-type' (e.g., observational versus interventional trials) and enforced rules that permit data collection to proceed when evidence of informed consent documentation is present.

In circumstances constituting public health emergencies, observational trials may be conducted using the criteria for non-research public health surveillance activities outlined in 45 CFR 46.102(l)(2). However, interventional trials imply significantly higher risk from administering an approved or experimental treatment outside of the standard of care, which means collecting informed consent is necessary. Current standard practice is to use a third-party vendor or partner (e.g., clinical research organization (CRO)) to collect and integrate eConsent into an integrated clinical trial platform. In the current state, we recommend rule-based configuration of these technical capabilities based on 'study-type' and restrict data collection lacking documented eConsent for interventional study-types.

Looking to the future, we recommend standardized software development to create an application-based consent option analogous to the one created by the NIH "All of Us" program that empowers: 1) a guided consent process for patients, 2) enables consent into new or additional trial activities, and 3) allows these individuals to revoke/end participation if they so choose.

7. User Interface and Experience

a. Optimizing the Experience of Health Care Professionals

To optimize the experience of health care professionals, we suggest:

- Requesting the smallest amount of information required to accomplish the goal.
- Streamlining and centralizing tooling. Limit the number of applications a user is required to train, log into, and interact with to accomplish a request. Whenever possible, utilize APIs and integration to limit duplicate data entry. Utilizing FHIR-based tools such as Questionnaire and CQL to identify data of interest and FHIR US Core-based APIs at the source can further minimize manual data collection. Additionally, using FHIR-based tools to disseminate guidance aiming to improve data quality, use of appropriate encoding, and data collection that can improve on completeness of data and maximizes automated retrieval versus manual submissions.
- Eliminating unnecessary actions. The Clinical One user interface (UI) is a modern design that eliminates unnecessary scrolling to the right to navigate to the next visit. Data entry is a continuous down a single page across forms within a visit and each is autosaved when moving from one to the next. Additionally, there is a single login/UI for all capabilities within the platform (e.g., design, Digital Gateway, randomization and trial supply management (RTSM), data collection, analytics, etc.).

b. Increasing Likelihood That Users Will Provide Input

To increase the likelihood that users will provide input, we suggest:

- Understanding users' time constraints and contexts by requesting the smallest amounts of information necessary to accomplish the goal.
- Using discrete, structured fields that ensure information is captured in a usable, normalized format that protects data quality.
- When possible, proactively displaying or pre-entering information using APIs and data integrations.
- Offering "super user" features, like keyboard shortcuts, to empower people and speed data entry.

The Clinical One UI design enables a more streamlined user experience with fewer clicks and allows more efficient data entry with a system designed to ease form and visit entry while respecting the users time with automated saves.

Clinical One also eases the burden of user management, including training for necessary capabilities. It has a streamlined user training within a single UI, where everything can be done all at once, minimizing the administrative burden. This frees up time and resources for the site staff, allowing them to focus on more valuable activities such as addressing queries faster and most importantly, patient care.

c. The Most Useful Existing Tools, Apps, and Processes

We recommend:

- Managing missing data by trying to understand the source of the missingness so that we can better understand the implications on the study design and results.
- Reviewing the attached example (please see Jaffe et al 2022 Validation of Race Cerner Real-World Data_RWD46.pdf) from a presentation at ISPOR, where missingness of race data was found to have the following potential impacts:
    - Selection bias: missing data were typically younger and female
    - Confounding bias: missing race may be associated with exposure and outcome
    - Measurement bias: code changes and non-standard entries could lead to potential for inaccurate association/inference – misleading results

8. Capturing Data Elements Required for Clinical Trial Protocols

To facilitate rapidly conducting clinical trials, we must increase the data collection bandwidth by opening new lanes for data capture and expanding “direct to consumer” channels for data capture. Such enhanced flexibility will allow us to leverage existing infrastructure and technology to expand the data capture process including (1) Use of Patient-Reported Outcomes Measurement Information System (PROMIS) API to deliver patient reported outcomes (electronic patient-reported outcomes (ePRO)) questionnaires directly to participants in accordance to a trials schedule; (2) Integration with APIs for prominent mobile data capture devices; (3) Ability to push electronic case report forms (eCRFs) directly to participants; and (4) Use of SMART on FHIR applications focused on providers to collect data that otherwise could not be automatically retrieved, bring all data into Clinical One, and facilitate interoperability.

Our goals include educating and empowering research participants via the decentralized clinical trials (DCT) functionality described above and the ability to change their informed consent status or opt-in/out for secondary use data sharing. We also seek to provide additional flexibility to researchers for the conduct of emergency clinical trials by providing participant communication and additional data capture options remotely, a critical feature in a pandemic afflicted world where travel may be curtailed. We posit that leveraging technology to educate and empower participants and research teams will foster collaboration and engagement fomenting trust and longitudinal partnerships in conducting research. Clinical One enables direct patient research and flexible scalable application of digital and decentralized study methods. The proposed will enable faster DCT study starts and provide sponsors and CROs choice in which DCT provider best meets their study needs. Optimizing direct patient data collection will optimize therapeutic and regional diversity, expand the scale of trials, and harbor all data in a consolidated source of truth and visibility regardless of who or where collected.

To ensure compliance with regulatory requirements, Clinical One is built to comply fully with multiple global regulatory requirements and guidelines, including 21 CRF Part 11, Electronic Records; Electronic Signatures - Scope and Application. This includes a fully validated system, supporting documentation for release testing, control of features to turn on, version control over the study design, full auditing for all changes, record copying and retention, and electronic signatures. Additionally, Clinical One supports fine grained role-based security control, single sign-on, and data blinding and unblinding by user access rights. Study designs are fully configurable and support reuse of existing design objects and additional event, form, and item controls can be configured to help guide users and query users when data appears to be invalid. For example, designers can add visit windows to the schedule of activities, restrict data types, code items based on dictionaries, and add code lists to questions with a controlled list of values. Further, simple validation checks can be configured such as dynamically displaying a visit, form, or question based on a response or specifying a valid range of values. More complex rules can also be added, such as cross-visit and form checking, calculations, and derivations.

9. TEFCA and QHINs

In the short- and medium-term, Trusted Exchange Framework and Common Agreement (TEFCA) adoption and compliance is voluntary for healthcare provider organizations while the research use case is not yet defined.

TEFCA already addresses two exchange purposes that should still be explored for the clinical trial data collection use cases under consideration: public health and individual access services. Public health or individual access services methods can be leveraged to obtain relevant clinical trial data to show how the clinical trial can be performed under a public health authority. It can also show how the clinical trial data collection approach uses clinical trial participant enrollment to obtain the relevant consent, which enables a consumer application to access the participant's own data managed by participating healthcare providers and share it with the clinical trial systems. At the same time, collecting clinical trial data should start to align with TEFCA such that once the Recognized Coordinating Entity (RCE) has introduced the research purpose, the transition is optimized. TEFCA is particularly relevant in collecting data from participating healthcare providers, although consideration should be given to how TEFCA's trust fabric can be utilized to streamline the adoption and use of clinical trial participant-focused applications.

Utilizing TEFCA at the earliest opportunity as described above allows relevant data to be gathered from the participating healthcare provider as well as any other participant-specific clinical context that is already available across all of the participant's clinical records. This would reduce both the burden on the healthcare provider who only needs to focus on new and unique data relevant to the clinical trial and the patient who does not need to provide that data again.

As individuals enroll in a clinical trial, a clinical trial participant will be provided with a clinical trial-specific application used to collect data best obtained directly from the participant themselves. Such interactions do not necessarily need to go through TEFCA's trust and exchange infrastructure. However, operating under TEFCA may

increase the trust relative to the actual data use and re-use provided in the clinical trial. As TEFCA initially focuses on document-based exchange and the documents available are typically not specific to any particular trial, challenges arise in the ease of using that initial method. However, we should consider the clinical trial-specific documents that participating healthcare providers can make available. That still poses challenges with obtaining other relevant data.

TEFCA's second phase, facilitated FHIR-based exchange, will provide the necessary tools to enable targeted, automated data collection in combination with complementary data collection by the healthcare provider where the data is not yet available or requires further curation before being submitted. The combination of the FHIR US Core, SMART Apps, as well as FHIR Questionnaire, Questionnaire, and CQL, within TEFCA's trust fabric enables the necessary scaling and flexibility to adjust, while focusing only on the relevant data. Establishing a research purpose of use standard operating procedure (SOP) is essential to determining how data collection guidance developed by the HL7 Vulcan accelerator would be adopted into the TEFCA fabric.

10. Emerging Technologies

Oracle is committed to leveraging emerging technologies in the healthcare industry. We envision a world where using the tools we have developed over decades of data management leadership and industry experience, combined with our comprehensive cloud applications suites, and our recent Cerner acquisition, allows us to make electronic healthcare records (eHR) patient-centric and not provider-centric. This patient-centric access to eHR would allow patients to access their records for visits in centers from California to Massachusetts while also enabling cloud-based access to anonymized data for emerging or pandemic-like scenarios.

Some areas of innovation within eHR aligns with patient anonymization of data and easier, more secure data sharing. Examples of emerging innovation areas include:

Tokenization

- This allows tokens to correspond to gene sequence, or mutation. By eliminating specific naming of those genes and traits, the data cannot be tracked by the investigator back to the individual without using a key that corresponds to the tokenization of the genes/traits.
- Datavant tokens are generated for every patient using identifiable patient demographic variables such as last name, first name, date of birth, etc. in an identifiable version of our Real World Dataset. The identifiable variables are then removed to ensure the Real World Dataset is HIPAA compliant according to two different de-identification methods (Expert Determination and Safe Harbor). Third party data providers also execute this tokenization process which then enables disparate datasets to be linked together while keeping the underlying patients de-identified.

Secured Container Model

- This model includes three parties: the owner of the data, the developer of the environment for artificial intelligence (AI), and a broker/trusted third party escrow for facilitation.
- Creation of a secure container model which hides data for AI/machine learning (ML) training, eliminating the user from observing any sensitive, protected medical information

To assist with reducing clinician work overload from data entry, areas within AI can be harnessed for innovation. AI powered voice to text applications are also being developed for use with Oracle Cerner's eHR to reduce the time site staff spend inputting relevant data. The use of AI and ML to facilitate patient care and trial efficiencies is a main area of focus for Oracle.

11. Pilot or Demonstration Project

A demonstration project is more appropriate than a pilot project for this initiative because the feasibility of centralized data collection and analysis for multi-site clinical trials is well established. A demonstration project would provide the resources to build operational infrastructure to facilitate readiness once emergency clinical trial services are needed. Using contemporary electronic data capture (EDC) solutions, data may be managed through a single central repository using cloud-based storage infrastructure. Cloud systems enable efficient provisioning of appropriate security protocols to protect patient-level data and redundancy to protect data integrity.

12. Specific Commercial Capabilities

When considering the implementation of emergency clinical trials, expediency is one of the most critical elements. To generate the speed required in situations where secondary data sources are not sufficient, we believe it is essential to shift from thinking in terms of individual studies to thinking in networks. Having established relationships with a coalition of research-ready sites that are trained and experienced allows for a low barrier to entry for participation in new trials.

Through standardization and automation of tasks across contracting, technology, data and site support operations, we believe we can transform former project-based, high-effort, low-value tasks into one-time efforts at a site level that can be leveraged with minimal effort on any new study. When the time comes for an emergency clinical trial to occur, this network will be ready to respond immediately.

Oracle is uniquely positioned to deliver on this network concept by leveraging our existing technology that spans the entire research lifecycle and developing new integrations to create a seamless end-to-end approach. We are working towards defining a strategy that can be implemented within the Learning Health Network, our opt-in model for health system clients interested in participating in research. Through the Learning Health Network, we can research-enable previously underrepresented health systems in the research community. This in turn improves access to a diverse set of patients, improving the validity and representativeness of clinical trial results.

Appendix A: General Terms and Conditions for Oracle's Proposal

Oracle America, Inc. ("Oracle") is pleased to have the opportunity to provide Office of Science and Technology Policy (OSTP) ("you") with a proposal for the RFI for a Data Collection for Emergency Clinical Trials and Interoperability Pilot Solution ("the response") released on October 28, 2022.

The response is subject to the following general terms and conditions:

The response is provided to you as confidential information and must be held in strict confidence. Your use of the information in the response shall be limited to your use solely in connection with evaluating the response. You may however share the response with your external advisors, agents and subcontractors on condition that the use remains limited to said purpose and subject to confidential treatment of the response; you are responsible for their confidential treatment of the response. You do not acquire any intellectual property rights in Oracle property under the response and you agree to comply with all applicable export control laws and regulations to ensure that no information is used or exported in violation of such laws and regulations. **If you do not agree with these terms, you are requested not to open the response and return it to Oracle as soon as possible.**

The response is based upon information that you have provided to Oracle and is intended for your evaluation purposes only. It is not for execution or incorporation into a contract that may result between you and Oracle. Neither you nor Oracle shall be obligated in any way until such time as we have agreed upon the terms and conditions and executed a final agreement.

Any Oracle program licenses ("programs"), hardware ("hardware"), technical support services, consulting services, software as a service or other services (collectively "services") will be provided in accordance with the terms of an Oracle Cloud Services Agreement ("Subscription Services Terms"), a sample of which has been provided, as may be amended by you and Oracle following award of the contract to Oracle (the "agreement") and one or more Oracle ordering document(s). Accordingly, Oracle takes exception to any provisions or requirements, which purport to establish any other terms and conditions for the provision of the Oracle programs, hardware and/or services. A future ordering document(s) will be governed by the terms and conditions of the relevant executed agreement. Such agreement and the ordering document(s) shall exclusively govern the terms and conditions under which the proposed programs, hardware and/or services will be provided.

The prices set forth in the response are exclusive of any sales, value-added or other similar taxes imposed by applicable law that Oracle must pay based on the programs and/or services, except for taxes based on Oracle's income. The prices are also exclusive of shipping and media charges. Shipping terms will be as specified in the agreement/ordering document. Documentation is provided in the form/format which is commercially available/industry standard for all customers. All fees payable to Oracle are due within 30 days from the invoice date.

If any future award includes programs, only those programs included in a response to a future proposal and included in a resultant contract shipment summary issued by Oracle in the applicable ordering document are available in production release on the computer hardware operating system combination(s) designated by you. Not all programs are available on all computer hardware/operating system combinations and Oracle is under no obligation to make available any program(s) or program/computer hardware/operating system combination except for the program(s) listed on a shipment summary issued by Oracle in an ordering document, executed by you and Oracle. Furthermore, the future availability of any program(s) or program/computer hardware/operating system combination shall not affect your payment obligations under any resultant agreement or relevant ordering document(s).

Unless agreed otherwise at the time of contracting, any Oracle consulting services are ordinarily provided on a time and materials basis.

Any future purchase of (a) hardware and/or related hardware support, (b) programs and/or related technical support, or (c) other services are all separate offers and separate from any other order for (i) hardware and/or related hardware support, (ii) programs and/or related technical support, or (iii) other services you may receive or have received from Oracle. You understand that you may purchase (x) hardware and/or related hardware support, (y) programs and/or related technical support, or (z) other services independently of any other product or service. Any future obligation to pay for (i) hardware and/or related hardware support is not contingent on performance of any other service or delivery of programs, (ii) programs and/or related technical support is not contingent on delivery of hardware or performance of any other service, or (iii) other services is not contingent on delivery of hardware, delivery of programs or performance of any additional/other service.

No statement made by Oracle in the response shall be construed as any representation or warranty including, but not limited to, implied warranties of fitness for a particular purpose, satisfactory quality or merchantability, representations or warranties as to performance, product or service availability, or any other representation or warranty and such provisions shall only be in accordance with the agreement and applicable ordering document(s).

Oracle assumes no responsibility for systems integration work or responsibility to act in the capacity as a prime or general contractor with respect to any third party products provided or services set forth in the response. Oracle makes no warranty as to the performance or suitability of any such third party products or services.

In the event of any inconsistencies between the text in other sections of Oracle's response and the text of this general terms and conditions document, the text of this document best clarifies Oracle's position and shall govern Oracle's entire response.

The information contained in this response is considered by Oracle to be proprietary and confidential to Oracle to the extent that such information is not currently available in the public domain. Subject to the Federal Freedom of Information Act 5 U.S.C. Section 552 (the "Act"), the information contained in this response may be used solely in connection with the evaluation of the response. To the extent that a claim is made to disclose confidential information contained in this response, Oracle reserves the right to defend its confidential information against such claim. You agree, subject to the Act, to (a) keep the information contained in this response in strict confidence and not to disclose it to any third party without Oracle's prior written consent and (b) your internal disclosure of the information contained in this response shall be only to those employees, contractors or agents having a need to know such information in connection with the evaluation of the response and only insofar as such persons are bound by nondisclosure obligations consistent with the foregoing. You do not acquire any intellectual property rights in Oracle's property under the response and you agree to comply with all applicable export control laws and regulations to ensure that no confidential information is used or exported in violation of such laws and regulations. You may make a reasonable number of copies of this response for your internal distribution for use solely in connection with the evaluation of the response; otherwise, you may not reproduce or transmit any part of this response in any form or by any means without the express written consent of Oracle. By reading the response, you have agreed to be bound by the foregoing terms.

Appendix B: Definitions and Abbreviations Used in Oracle's Proposal

Term	What It Means
achieve	Oracle and our clients benefit when we agree in writing to a set of standards for objective performance and intellectual property. When this phrase is used, it intends to mean that Oracle will comply with obligations that are codified in contracts with our clients.
AI	Artificial intelligence
API	Application programming interface
certify or certified	When this term is used in regard to an Oracle product, it is not intended to imply compatibility of products or that products have been tested to address particular business requirements. When this term is used in regard to an Oracle employee, it reflects Oracle's view of the experience we can provide and/or an individual's personal development through Oracle-recognized programs. In all cases, Oracle will comply with obligations that are codified in contracts with our clients.
CDS	Clinical decision support
configure or configuration	The setup of the applications by entering specific values which drive business processes using the Standard Functionality provided within the Oracle application(s) without extension.
CQL	Cassandra Query Language
CRF	Case report form
CRO	Clinical research organization
DEQM	Data Exchange for Quality Measure
develop or development	Oracle is, in part, a software development company. When we use the word "develop" or its derivatives outside of the context of how Oracle has built our standard suite of products, "develop" or its derivatives intend to mean that Oracle will comply with obligations that are codified in contracts with our clients.
DMW	Oracle Data Management Workbench
eCRF	Electronic case report form
EDC	Electronic data capture
eHR	Electronic healthcare record
enhance	When "enhance" is used in context of augmenting a product's or system's performance, Oracle means that the solution described in our proposal is believed to be able to assist you with addressing the business issues outlined in the RFX. Oracle does not use this word to imply a guarantee or warranty. Oracle will comply with obligations that are codified in contracts with our clients.
ensure	Oracle and our clients benefit when we agree in writing to a set of objective performance and delivery standards. When this phrase is used, it intends to mean that Oracle will comply with obligations that are codified in contracts with our clients.

Term	What It Means
ePRO	Electronic patient-reported outcome
expert or expertise	Our customers benefit from Oracle Health Sciences' vast experience and dedication to the industry. When these terms are used, they reflect Oracle's view of the experience we can provide. Note that Oracle will comply with obligations that are codified in contracts with our clients.
FHIR	Fast Healthcare Interoperability Resources
HIPAA	The Health Insurance Portability and Accountability Act
HL7	Health Level Seven
HSC	Oracle Health Sciences Consulting
improve, improvement	When "improve" or "improvement" is used in context of augmenting a product's or system's performance, Oracle means that the solution described in our proposal is believed to be able to assist you with addressing the business issues outlined in the RFX. Oracle does not use this word to imply a guarantee or warranty. Oracle will comply with obligations that are codified in contracts with our clients.
integration or integrate	Except to the extent expressly stated in the scope section of this document, the use of the terms "integrate" and "integration" throughout this document is not intended to mean that Oracle will address (i) the physical or functional integration of Oracle products with external legacy applications, third-party products, and/or other software applications; (ii) the functioning of Oracle products as a coordinated whole with such external legacy applications, third-party products, and/or other software applications; or (iii) any non-standard integration between Oracle products. Rather, the terms are used to refer to the overall concept of data exchange between the Oracle products and other applications, products, or applications identified in this document, and may include interfacing and/or other methods of integration or interoperation as described in the scope section of this document.
meet or exceed your needs, requirements, expectations, or similar	Oracle and our clients benefit when we agree in writing to a set of objective performance and delivery standards. When these phrases are used, they intend to mean that Oracle will comply with obligations that are codified in contracts with our clients.
ML	Machine learning
NIH	National Institutes of Health
ONC	The Office of the National Coordinator for Health Information Technology
Oracle	Oracle Corporation
out of the box	This phrase is not meant to imply that a product will meet a customer's business needs (or expectations) without any special configuration or customization. Instead, these are used to reflect the product's standard functionality. "Standard functionality" for an application is defined as the functionality described in applicable documentation for the application as provided by Oracle.
partner or partnership	The term "partner" or "partnership" refers to and is interchangeable with "ally" or "collaborator". Use of the term is not intended to, and does not, contractually or otherwise bind Oracle to the client, or create a partnership, joint venture or agency relationship between Oracle and the client.

Term	What It Means
PROMIS	Patient-Reported Outcomes Measurement Information System
RCE	Recognized Coordinating Entity
RTSM	Randomization and trial supply management
SANER	HL7 Situational Awareness for Novel Epidemic Response
solution	The term “solution” is not intended to, and does not, express or imply that Oracle can or will contractually or otherwise agree to “solve” any issues or problems.
SOP	Standard operating procedure
SSN	Social security number
support, supported, or not supported	“Support” and its derivatives have many meanings. “Supported” sometimes refers to whether a program is covered under a contract for technical support. In addition, "supported" may refer to whether a certain business process may be addressed using functionality contained in a standard product configuration. “Supported” may also be used to identify products or features that work together or are compatible. MyOracle Support provides technical assistance for Oracle customers. Oracle leadership lends their “support” to our teams on the ground. Because this RFX seeks information on a number of types of support, we drew heavily on context to create answers to the questions asked. Oracle will comply with obligations that are codified in contracts with our clients.
system	When Oracle uses the word “system,” we mean it to be a “platform” or “environment.” The use of the word "system" does not extend to Oracle any responsibilities to third-party components, systems, and/or products that you are responsible for when Oracle is only delivering our products or Cloud services. Oracle will comply with obligations that are codified in contracts with our clients.
TEFCA	Trusted Exchange Framework and Common Agreement
UI	User interface
US	United States
USCDI	United States Core Data for Interoperability

January 24, 2023

The Honorable Arati Prabhakar
Director
White House Office of Science and Technology Policy
1650 Pennsylvania Avenue
Washington, D.C. 20504

RE: Notice of Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Dear Director Prabhakar:

The Healthcare Leadership Council (HLC) appreciates the opportunity to submit comments on the White House Office of Science and Technology Policy's (OSTP) request for information on, "Data Collection for Emergency Clinical Trials and Interoperability Pilot."

HLC is a coalition of chief executives from all disciplines within American healthcare. It is the exclusive forum for the nation's healthcare leaders to jointly develop policies, plans, and programs to achieve their vision of a 21st century healthcare system that makes affordable high-quality care accessible to all Americans. Members of HLC – hospitals, academic health centers, health plans, pharmaceutical companies, medical device manufacturers, laboratories, biotech firms, health product distributors, post-acute care providers, home care providers, and information technology companies – advocate for measures to increase the quality and efficiency of healthcare through a patient-centered approach.

The COVID-19 public health emergency (PHE) has highlighted the enormous role that clinical research and data play in responding to disaster events. Public and private stakeholders need to be able to quickly collaborate to advance medical countermeasures from therapeutics to vaccines as well as to share information with each other to make necessary decisions about readiness and response efforts. In February 2021, HLC in partnership with the Duke-Margolis Center for Health Policy published a report on how to improve our nation's disaster readiness infrastructure. HLC supports the work of the Accelerating COVID-19 Therapeutic Interventions and Vaccine (ACTIV) partnership as a blueprint for bringing together diverse stakeholders. Future partnerships should build upon this framework of bringing public and private groups together while streamlining regulatory approval.

One of the greatest barriers to swift transmission of information to support supply chains, care delivery response, and biomedical innovation is the lack of a recognized or utilized single data standard to allow systems to be interoperable with one another. Use of a recognized, single data standard for information exchange allows for timely use of decentralized data (and the use of data not previously anticipated as necessary) to support all the components of response from supply chain management to outbreak tracking to clinical research essential to medical

countermeasures. We encourage you to examine steps to transition to an industry-supported data standard, such as HL7 FHIR, to enable stakeholders to provide necessary information in a timely manner to improve readiness and response efforts, support decentralized clinical research, and allow for timely and flexible data analysis integral to readiness.

Additionally, HLC thanks the Office of the National Coordinator (ONC) for its work to advance the Trusted Exchange Framework and Common Agreement (TEFCA). Implementation of this framework will help to establish a floor of universal interoperability and connectivity to enable efficient and secure access to and sharing of health data. This will improve data availability at the point of care and allow patients to become more involved in decisions about their care. The Centers for Medicare and Medicaid Services (CMS) is currently examining how to integrate TEFCA into CMS programs to enable providers to share information among one another. We encourage OSTP and ONC to examine how to further leverage TEFCA for public health data use cases, both for clinical trial research and disaster readiness and response.

HLC looks forward to working with you to improve data collection efforts. Please contact Tina Grande at (202) 449-3433 or tgrande@hlc.org with any questions.

Sincerely,

Mary R. Grealy
President

Response to RFI: Clinical Research Infrastructure and Emergency Clinical Trials
Document Citation: 87 FR 64821

David S Stephens MD* and Kathleen M. Neuzil, MD, MPH**

*Vice President for Research

Robert W Woodruff Health Sciences Center

Stephen W Schwarzmann Distinguished Professor and

Chair Department of Medicine

Emory University

**Myron M. Levine MD, DTPH Professor in Vaccinology

Professor, Medicine and Pediatrics

Director, Center for Vaccine Development and Global Health

University of Maryland School of Medicine

We [National Institute of Allergy and Infectious Diseases (NIAID) sponsored Infectious Diseases Clinical Research Consortium (IDCRC)] representing the nation's ten Vaccine Trials Evaluation Units, (VTEUs) https://idcrc.org/] are writing in response to this RFI based on our engagement and leadership (2020-present) in the pivotal COVID-19 prevention (e.g. vaccine, mAb) and therapeutic (small molecule/drug) clinical trials. The extramural IDCRC and VTEUs work in tandem with NIAID and other federal agencies as a coordinated national and global network of scientific experts to develop and test vaccines and other therapies to combat infectious diseases. In early 2020, we were mobilized to plan and successfully conduct at IDCRC sites the first phase 1 mRNA (Moderna) vaccine trial (began 65 days after sequence of the virus available). The data on this trial were collected and published on July 14th, 2020. Subsequently, the IDCRC, as part of the COVID-19 Prevention Network (CoVPN), was instrumental in the design, conduct and leadership of the five large (136,000 participants) Phase 3 trials of COVID-19 vaccines leading to multiple FDA vaccine authorizations and approvals, one of the first (the Moderna mRNA vaccine) was authorized in Dec. 2020. VTEU investigators were co-principal investigators on each of these trials. Other ongoing IDCRC COVID-19 trials (2020-present) include the mRNA vaccine boost and variant studies, "Mix and Match" of different COVID-19 vaccines, pediatric (Kid-Cove) and pregnancy studies (MOMI-Vax) of COVID-19 vaccines, and leadership in the COVID-19 vaccine Variant Immunologic Landscape Trial (COVAIL). We also were a leading network in the Adaptive COVID-19 Treatment Trials (ACTT1-4) the latter showing the value of

remdesivir, baricitinib and corticosteroids. These trials have significantly influenced US and international public health policy, and secondary analysis of data, including correlates analyses, continue to inform updated policy recommendations and approvals of secondary generation vaccines.

IDCRC does not agree with elements of the first basic premise in the RFI:

"The lack of a coordinated approach to clinical trials research in emergency settings has slowed the development of actionable information, which has in turn delayed the availability of vaccines, therapeutics, and diagnostics; and may also impede the tracking of the outbreaks themselves. Without some mechanism to coordinate and organize research on a larger scale in an emergency setting, researchers and decision makers are left with a series of relatively small, often inconclusive studies, and assembling data for larger-scale analysis is challenging."

There was a rapid mobilization of our network first by NIAID and subsequently by NIH and the public-private partnership Operation Warp Speed (OWS) now the USG Countermeasures Acceleration Group and the White House COVID-19 Response Team. The studies of the IDCRC, ACTT Consortium and CoVPN conducted in this public health emergency both in the US and at global sites were not only rapid but of very high quality and incredibly impactful (https://pubmed.ncbi.nlm.nih.gov/36689221/) . The studies demonstrated safety and efficacy of new vaccines and vaccine technologies, established successful therapeutics, and have saved millions of lives. As an example, these studies contributed to our global vaccine leadership that the US continues to enjoy, have led to effective treatment options for COVID-19, resulted in the effective vaccination of US and global populations, and identified vaccine and therapeutic products that should not be pursued. A similar success story can be said of the NIH RADx initiative in advancing new COVID-19 diagnostic technologies.

Decisions that impact public health must be built on rigorous scientific data. Product selection for trials must have a solid scientific basis and demonstrate safety in preclinical studies. Products with faulty design or poorly designed or executed clinical trials will not give clear answers and may in the end be harmful. While health care providers must be engaged in understating the value of clinical trials, most are not trained in the science or rigor of clinical research, and crucial regulatory (e.g., human subjects protection and informed consent, documentation) requirements for such research. We cannot rely on anecdotes, cases series, observational studies, or "pragmatic" studies to substitute for rigorous clinical trials. Hydroxychloroquine is a key example of this point. Bottom line we need to have a national infrastructure supporting training of clinical trialists and an infrastructure ready to respond to national emergencies.

The current U.S.-level governance structure: HHS ASPR coordinating with NIH, BARDA, DOD, FDA can be improved and streamlined but is an appropriate governance structure for coordinating the US approach to clinical trials research in emergency

settings. We do strongly support the efforts to improve and modernize electronic data entry and clinical trial data collection across trial sites that can be scaled up for use in emergency research settings but under an umbrella of a rigorous clinical trials infrastructure.

IDCRC strongly agrees with a second premise in the RFI

"a key issue is to support the expansion of clinical research into underserved communities and increase diversity among both trial participants and clinical trial investigators"

Both increased diversity in trial participants and diversity in clinical trials investigators need better planning and additional governmental leadership and resources. An example of an effective approach to educate underserved communities in clinical research was developed by the CoVPN for the phase 3 COVID-19 vaccine trials and included the CoVPN Community and Stakeholder Engagement Strategic Plan (attached) and launched the related CoVPN Faith Initiative https://www.coronaviruspreventionnetwork.org/about-covpn/. The enrollment at CoVPN sites of underrepresented minorities in the Phase 3 vaccine clinical trials was exceptional. (https://pubmed.ncbi.nlm.nih.gov/36689221/)

The second component "enhancing diversity among investigators" is also strongly endorsed. We recognize the need for formal training in the discipline of clinical research and vaccinology has never been greater. We are positioned to equip a new generation of scientists with the necessary tools to enable them to explore, create, innovate and implement the vaccine and treatment programs of the future. As successful examples we highlight the IDCRC Mentorship Program https://idcrc.org/training/index.html and the Early Career Investigator Pilot Awards https://idcrc.org/training/pilot-grants-program.html providing mentorship, professional development and funding of early career investigators and fellows in clinical and translational infectious diseases research. Also, the CTSA infrastructure supporting education (Master's level degrees) and training in clinical and translational research is another example. Our program continues to innovate and incorporate new technologies, strategies, data analytic tools and educational approaches to prepare the next generation of leaders in clinical and translational research.

In summary, we wholly endorse the need for clinical trial infrastructure, and put forth the IDCRC and VTEUs as an example of how investments in time, talent, leadership and infrastructure were key to the rapid and successful COVID-19 response in the U.S. We endorse using this already strong program as a foundation from which to build a more robust and diverse emergency response infrastructure in the US.

Dear OSTP and ONC,

On behalf of the Faro Health team, I am pleased to provide information to support your RFI for Data Collection for Emergency Clinical Trials and Interoperability Pilot. As our CEO and co-founder of Faro Health, Scott Chetham, vocalized during your panel discussion January 11th, the actual clinical trial design is critical to standardizing data collection. The clinical protocol truly serves as a blueprint as to how the trial is to be conducted and which patient population it targets. Without a well-thought blueprint (protocol design), you increase your risk for high patient attrition rates, increased patient burden and potential time delays and costs due to inefficiencies not identified early enough. As an industry, we must take a step back and learn from the past and work to get the protocol design right-sized initially and not just reactively enable methods of risk management and data collection in a fast and furious way. We can actually get ahead of this process and eliminate numerous errors and obstacles along the way if the design of the protocol is collaborative and efficient. This is where Faro Health comes into play with our Faro Smart Designer.

Faro Health brings life science professionals together in a user-friendly environment, empowering multiple teams to collaborate and balance the complexity of modern trial designs and see the impact of their decisions on patients and sites all in real-time. In emergencies clinical protocols must be clear, concise and accessible. We must have clear and precise data models in order to readily aggregate protocol data. The Faro Smart Designer promotes data collection standardization by embedding industry-vetted standards throughout the protocol, most notably in the schedule of activities. Faro follows key principles of Transcelerate's digital data flow initiative to enable seamless flow of standard assessment/measurement metadata to automate the programming of downstream clinical data capture tools. Digital clinical protocols enable sharing and collaboration of complex information that is not possible using Microsoft Word and PDFs.

Faro Health
Jan. 25,2023

As clinical trials become more and more complex, it is critical for efficiencies and modifications to be identified to decrease burden and avoid cumbersome implementation. Our platform, which is powered by standards, is intuitive as it makes recommendations for decentralized approaches while allowing stakeholders to see the impact of their decisions. Critical intel such as patient and / or site burden, cost impact and even blood volume are available as the schedule of activities is being built for the protocol. This allows for greater efficiency of protocol implementation and data collection.

Faro Health would be thrilled to participate in initiatives led by OSTP and ONC to help streamline the critical design of the protocol for upcoming health needs. I have provided below some information about Faro along with an appendix of screenshots from our platform to show just how powerful this platform can be.

Sincerely,

Kimberly Pospahala
VP of Commercial
Faro Health
kimberly.pospahala@farohealth.com

I. About Faro

Faro Health is bringing clinical trials into the digital age by helping teams manage and balance the complexity of modern trial designs through a cloud native platform. The Faro platform enables study teams to design complex clinical trials using small modular building blocks and combines that with data driven insights to orchestrate and automate operationally complex trials. Faro brings balance, centricity and flexibility to protocol development through automation and integration to downstream systems and vendors, ensuring they are always up to date and correctly configured. As a result, clinical trials using Faro are operationally efficient which avoid delays due to ambiguity and generate data that can be trusted.

II. Description of Products

Faro Health's Study Designer and platform addresses the limitations and errors caused by the manual, document and spreadsheet-driven asset creation and maintenance processes. Complex studies can now be assembled from modular components that can be used to automate the creation of trial assets. Real Time insights into the trial design can help teams visualize the budget and patient burden and inform critical decisions. Multiple trial design scenarios can be created to help teams answer the right question in the least burdensome fashion for patients, sites and regulators. Integration with downstream clinical systems and vendors (Electronic Data Capture Systems, Clinical Trial Management Systems) offers the opportunity to simplify and automate previously manual labor-intensive tasks. Structured data and content can be repurposed ensuring traceability and eliminates the need for time-consuming third-party conversion. This ensures a fully traceable chain of custody for the trial data as the study design is operationalized.

III. Faro Smart Designer Platform Examples

Faro Smart Designer

SaaS platform that eliminates manual process and leverages technology to assist with lean protocol design and authoring.

Built by designers for designers

Ideas & RWD
Combine study concepts and RWD to see the reality of patient impact

Less is More
Helps teams to balance scientific complexity.

Operational Reality
Analyze trial designs for operational feasibility.

Flexibility is Critical
Allows design teams to work without rigidity in process while maintaining structure.

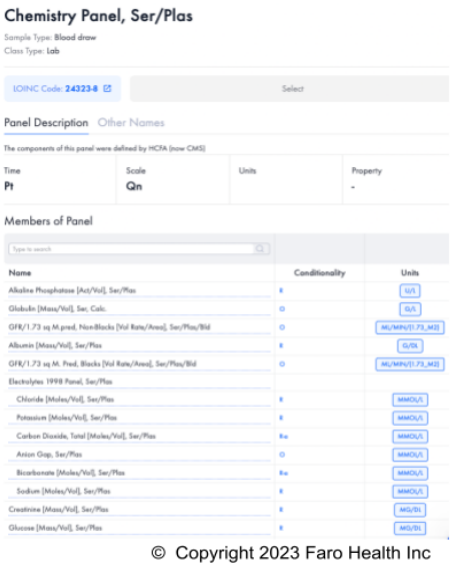
Idea to Reality
Push trial designs into workable documents.

Digital Data Flow
Prepare for downstream systems integrations with an accurate single source of truth.

faro

Linked to Global Lab Standards

- Ensure the proper assessments and measurements are defined without variability
- Maintain continuity across protocol designers to ensure accuracy and streamline the build process

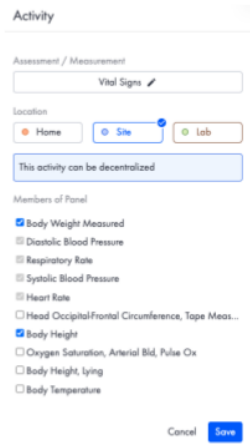


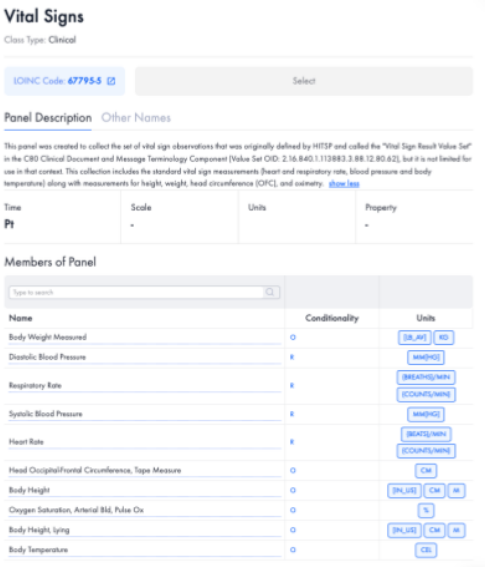
faro

Faro Health
Jan. 25,2023

Standard Metadata Definitions

Standard assessment and measurement metadata is populated to enforce consistent data collection





faro

Optimize Study Designs With Real Word Data

Compare two different design scenarios with a click

Study activity costs with Liver biopsy

Study activity costs without Liver biopsy

faro

Faro Health
Jan. 25,2023

Patient Journey

Teams can use real world data to view the study from different perspectives

Site Perspective

Teams can use real world data to view the study from different perspectives

Faro Health
Jan. 25,2023

MITRE's Response to the OSTP RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot

January 26, 2023

For additional information about this response, please contact:

Duane Blackburn
Center for Data-Driven Policy
The MITRE Corporation
7596 Colshire Drive
McLean, VA 22102-7539

policy@mitre.org
(434) 964-5023

About MITRE

MITRE is a not-for-profit company that works in the public interest to tackle difficult problems that challenge the safety, stability, security, and well-being of our nation. We operate multiple federally funded research and development centers (FFRDCs), participate in public-private partnerships across national security and civilian agency missions, and maintain an independent technology research program in areas such as artificial intelligence, intuitive data science, quantum information science, health informatics, policy and economic expertise, trustworthy autonomy, cyber threat sharing, and cyber resilience. MITRE's 10,000-plus employees work in the public interest to solve problems for a safer world, with scientific integrity being fundamental to our existence. We are prohibited from lobbying, do not develop or sell products, have no owners or shareholders, and do not compete with industry. Our multidisciplinary teams (including engineers, scientists, data analysts, organizational change specialists, policy professionals, and more) are thus free to dig into problems from all angles, with no political or commercial pressures to influence our decision-making, technical findings, or policy recommendations.

MITRE has carried out several efforts investigating methods of data extraction, transformation, gathering, analysis, and interpretation in support of multiple federal agencies. These activities span the breadth of FFRDCs that MITRE operates, including non-health domains such as defense, cybersecurity, and intelligence. We have conducted investigations and pilots in the realms of novel clinical terminology and transport standards, methods of privacy preserving record linkage, federated learning networks, and the use of Fast Healthcare Interoperability Resources (FHIR) at scale for clinical research. Data and insights from these activities form the basis to this response.

Introduction and Overarching Recommendations

MITRE's recommendation for emergency clinical trials is a system for routine clinical trials that is regularly exercised, routinely improved upon, and available for emergency use when needed. Such a network would ideally have several features:

1. The network should be a pragmatic one, comprised of data available as a consequence of routine clinical care. Note that this does not exclude interventional trials. During our participation in the COVID-19 Healthcare Coalition and with our health system partners, we demonstrated the use of a pragmatic approach to interventional trial designs with data submission via a common data model.[1]
2. The network should have a routine care use that encourages its regular exercise and maintenance. Some possibilities include routine reporting to registries, post-market pharmaceutical surveillance, or monitoring of pharmacologics under a Risk Evaluation and Mitigation Strategy (REMS).

[1] COVID-19 Healthcare Coalition. 2022. MITRE, https://c19hcc.org/. Last accessed January 23, 2023.

3. The network should be available to anyone. Open science has the potential of making the scientific process more transparent, inclusive, and democratic.[2] Citizen scientists and researchers at small organizations have equal potential to contribute to growth of medical knowledge, and the means for rapidly evaluating large corpuses of data are now generally available. Furthermore, the availability of data means that anyone can replicate a study design and independently validate a conclusion. In emergency clinical situations, open access to data means that individual scientists can continue to evaluate potential treatment options as more clinical experience becomes available.
4. The network should have provisions for emergency use, potentially including a lightweight data use agreement for emergency purposes. MITRE demonstrated the use of such an agreement in the COVID-19 Healthcare Coalition, and we were successful in getting over 900 organizations to accept the data use agreement. An alternative is to have emergency use as part of the data use agreement for the network, with clear indications as to what circumstances the emergency use authorization would be used under.
5. The network should have provisions for deidentified and identified research. Participant organizations should have the ability to locally define the queries to which they will respond. These definitions can be built into a firewall around an organizations' data, and only queries that meet the organization's acceptable use should be permitted through. Open source and commercial solutions for such networks exist, and several networks have opened in Europe using them. MITRE has demonstrated the viability of these networks in the lab using synthetic clinical data.
6. The network should allow for patients to conduct operations on their own data.
    a. Patients should be able to view their data in the network at any time.
    b. Patients should be able to opt-out of routine, de-identified data sharing at any time unless the network is used for a purpose that is exempt, such as quality assurance or use as part of payor operations.
    c. Patients should be able to opt-in to sharing their data for targeted research purposes, such as rare disease registries or emergency observational trials that may require protected health information (PHI).
7. Data made available to the network should be delivered using a common data model. The Observational Medical Outcomes Partnership (OMOP) common data model, for example, has been shown to rapidly accelerate the development and execution of research efforts. The use of FHIR for research is complex as its use is transactional in nature. We do note that research to derive OMOP from FHIR has grown positively in the last few years. Efforts to invest in and derive a common data model from FHIR may prove critical to health system participation, especially for resource-constrained health systems that may need the ability to develop and maintain their own common data model extraction, transformation, and loading (ETL) process.
8. Policy development may be necessary to expand constructs such as "safe harbor" to programmatically protected data assets. For example, if a health system's data store is available to the network and tested to be conformant to a given definition of privacy preservation, then responses to requests for de-identified data via such mechanisms

[2] Open Science. 2023. UNESCO, https://www.unesco.org/en/open-science. Last accessed January 20, 2023.

should not be considered PHI. This will further support the participation of under resourced hospitals and community health centers that may be unable to support expert review of every query on the network for risk of PHI disclosure.

Questions Posed in the RFI

Due to space limitations, MITRE has chosen to answer the first six questions posed in the RFI. Our responses to these questions briefly touch on points we would have made on the remaining questions if space was available, and we welcome the opportunity to expound further if desired.

1. **United States Core Data for Interoperability (USCDI).** We seek input on how U.S. Government and external stakeholders might leverage USCDI and future extensions of USCDI standards (such as USCDI+, an extension that supports federal partner program-specific requirements) to support emergency clinical trial research. It would also be helpful to receive comment on areas in which additional extensions might be necessary.

MITRE recommends that the data needed to support emergency clinical trial research should be piloted, prototyped, and moved into the relevant policy and regulatory constructs prior to need arising. The data should be conformant to a common data model to facilitate the development of study designs and protocols that can be universally implemented. Such data should derive from routine care provision and be used for the monitoring of routine care as frequent use will ensure that the data is available for emergency purposes when needed.

Principally, the standards and infrastructure necessary for emergency clinical trial research are not different from the ones that could support routine clinical trials. With MITRE's support for the mCODE effort and the Health Level Seven International (HL7) CodeX FHIR Accelerator community, we have learned that FHIR can be used for data collection in support of clinical trials. The Alliance for Clinical Trials in Oncology has leveraged mCODE for the ICAREdata project, a research effort to demonstrate that pragmatically collected data made available using FHIR is equivalent to traditionally collected data for research trials. In phase one, data was automatically extracted via language processing from clinical notes and compared to the gold standard. In phase two, data was captured through the use of routine clinical documentation tools and delivered via FHIR. To this end, MITRE and the Alliance for Clinical Trials in Oncology partnered with Epic Systems Corporation to develop and deploy the ICAREdata documentation tools. These tools, and the ability to capture ICAREdata, have been available to any Epic customer for several years.

Since implementing ICAREdata, several organizations have adopted the ICAREdata questions across their entire relevant patient population. They chose to do this for a variety of reasons, including but not limited to standardizing clinical process, acquiring more data to improve their own operations, and use of mCODE beyond clinical research. The CodeX community is exploring the use of mCODE for clinical trial matching, registry reporting, REMS reporting, quality measurement, and other use cases.[3] The use of mCODE for routine care means that the data is available when needed for many purposes, including clinical research.

[3] CodeX Use Cases. 2023. Confluence, https://confluence.hl7.org/display/COD/CodeX+Use+Cases. Last accessed January 20, 2023.

The use of pragmatically collected data for research purposes was also demonstrated during the COVID-19 pandemic. Below are some examples:

- The COVID-19 Healthcare Coalition built rapid clinical studies based on common data models and collection techniques. Through these efforts the coalition was able to answer key clinical questions quickly and effectively on the use of targeted therapeutics in the early days of the pandemic. In areas where the needed data was not part of the common data model, the coalition was able to rapidly prototype expansions of the common data model to support these studies. The coalition then worked with a few health systems that were able to include these expansions as part of their processes, leading to the generation of preliminary results. These interim efforts allow for preliminary results while the common data model and network are updated, as well as for efforts where updating the entire network is not needed or warranted.
- The Observational Health Data Sciences and Informatics (OHDSI) community was able to leverage its existing common data model and collaborator network to answer many questions during a three-day virtual connectathon, just a few months after the discovery of SARS-CoV2. Because their research capabilities were already in use, this community was able to complete clinical trials that others took months or years to replicate.[4]
- The theoretical utility of FHIR for epidemic response was demonstrated by the Situational Awareness for Novel Epidemic Response IG and its team.[5]

MITRE's efforts with mCODE have demonstrated that there is a role to expand U.S. Core (and USCDI) to support the data elements needed for clinical research. While mCODE derives from U.S. Core, it is an expansion on U.S. Core to fulfill needs for oncology purposes. The CardX and GenomeX communities are now developing similar solutions for the cardiology and genomics domains. A similar expansion for infectious diseases would not only support pandemic use cases but could also provide novel and real-time insights into the progression of influenza, streptococcal pneumonia, MRSA, or other topically relevant pathogens. In such an effort, MITRE would urge that the notion of a "minimal" common set of data elements (such as mCODE for oncology and mCARD for cardiology) be used as the grounding for the effort. Perfect is the enemy of good, and efforts such as the OHDSI community's COVID-19 connectathon showed that a significant number of very important use cases can be asked and answered of "good" data.

2. **HL7 FHIR APIs**. We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-emergency phase, and in what areas further advances might be needed.

MITRE recommends that the functions for data movement in a pilot research network follow the framework laid out by DaVinci Data Exchange for Quality Measures (DEQM). While not every operation in DEQM may be necessary, standards such as Bulk FHIR and FHIR Questionnaire

[4] 88 Hours: OHDSI's Signature Moment. 2023. Observational Health Data Sciences and Informatics, https://www.ohdsi.org/88-hours/. Last accessed January 20, 2023.

[5] Situational Awareness for Novel Epidemic Response. 2022. HL7 International, https://build.fhir.org/ig/HL7/fhir-saner/. Last accessed January 20, 2023.

are relatively more mature and should be strongly considered for the pilot. When FHIR resources are not natively available, consider the deployment of lightweight helper applications such as the mCODE Extraction Framework to facilitate piloting the remainder of the network. Finally, given the significant efforts that have been spent developing and refining OMOP CDM and its use for clinical research, we recommend evaluation of hybrid approaches such as OMOP-on-FHIR as part of the pilot.

In the DEQM, the DaVinci team developed three core use cases for data movement:

- An ongoing "reporting" mode for low frequency, "should never happen" events (such as catheter-related blood or urinary tract infections)
- A frequent "low volume" exchange mode for targeted quality improvement campaigns (such as Million Hearts or Accountable Care Organization related targeted campaigns)
- An infrequent "high volume" exchange mode for population level reports (such as hypertension or smoking cessation metrics).

The clinical trials that ICAREdata is supporting are "low volume" use cases. The MITRE mCODE team elected to use FHIR Messaging as our data movement method. We developed an open source, freely available utility that allowed health systems to export CSVs of data locally and transmit them using FHIR Messaging.[6] We developed a native FHIR application as well, but this utility has not been implemented to date for several reasons:

- mCODE APIs are not readily available.
- To work around the lack of native mCODE APIs, our FHIR application used proprietary APIs. These limit sharing and use of the application outside of the ecosystem it was developed for.
- Health system CIOs and CISOs were reticent to share data via API due to lack of sufficient local auditing and security control resources.
- Health system CIOs and CISOs are very comfortable with CSV extracts, which they have been supporting for decades.

In our discussions with the electronic medical record vendor community, the use of helper applications such as our mCODE Extraction Framework have been welcomed. These allow for prototyping of FHIR ecosystems without vendors having to invest in exposing APIs for relatively immature standards. The mCODE Extraction Framework was also welcomed by CISOs and CIOs. MITRE recommends the use of helper applications be strongly considered when piloting or prototyping a novel use case or standard. These allow for pilots and prototypes to be conducted in a more cost-effective way, facilitating good participation in the piloting phase. When standards such as mCODE are considered for further movement along the policy process and ultimately in the inclusions of all EMRs, the community can be confident the standard being adopted has been well exercised.

For pragmatic clinical trials, or ones where novel data gathering is required, FHIR Questionnaire is a potential way to deliver the query. MITRE chose another approach with the mCODE effort, primarily as FHIR Questionnaires were not available at our development partners at the time of

[6] A Node.js framework for extracting mCODE FHIR resources. 2023. GitHub, https://github.com/mcode/mcode-extraction-framework. Last accessed: January 20, 2023.

designing the ICAREdata study. We collaborated with clinicians on the data elements they most needed for oncology purposes and with Epic Systems Corporation on how they wanted to present the questions. Ultimately, this led to Epic developing a common (or Foundation System) form for collecting mCODE, which they made available to all their customers. Health systems were also free to incorporate those questions into their own forms or design their own questions. These implementations were reviewed by the ICAREdata team to assure they met the requirements of the clinical trials being supported.

One consideration regarding the use of APIs for research purposes is the nature of APIs. FHIR assets made available by APIs are intended to be use for data transactions. Use of routine FHIR APIs for high volumes of data would result in a tremendous number of data calls. Bulk FHIR is a relatively mature standard that facilitates such operations over groups or all patients in a system. In the relevant CodeX use cases such as registry reporting, our use cases have opted to use other mechanisms due to the availability of Bulk FHIR at pilot sites, the suitability of other FHIR operations for the use case, or the familiarity of partners with other FHIR operations and utilities. We continue to support the use of Bulk FHIR in DaVinci and look forward to being able to leverage it in further prototypes, pilots, and implementations.

The use of FHIR APIs as an indirect means for supporting clinical research should also be considered. The Observational Health Data Sciences and Informatics (OHDSI) community has been supporting clinical research since 2014, building upon the OMOP common data model from previous efforts. OHDSI has also developed a host of open source, freely available tools supporting the development and execution of clinical research targeting the OMOP CDM.

MITRE notes that research into deriving OMOP CDM from FHIR resources has greatly expanded in the last few years. In 2022, on behalf of (HL7) MITRE attempted to develop an environment for developing quality measures that could target both OMOP CDM and FHIR endpoints.[7] Many of the capabilities MITRE leveraged in that effort were limited prototypes or proofs of concept. As a result, MITRE was able to demonstrate the entire use case at the 2022 September HL7 Connectathon, but the real-world utility of such efforts is limited by the scope of the underlying capabilities.

We are confident that, with additional research and support, it should be possible to derive data conformant to the OMOP CDM from FHIR resources in the future. This future would have several advantages:

- Health systems and EMR vendors could focus on the provision of data via FHIR APIs.
- Researchers and other interested parties could develop studies using the OMOP CDM and the tools that already exist.
- Code for such studies could be shared using the same mechanisms the OHDSI community does today, collaboratively refined upon, and executed by any party with access to a clinical data network.

[7] Reference implementation software and sample data for supports testing OMOP to FHIR-based transformations, with an initial focus on Digital Quality Measures (dQM). 2023. GitHub, https://github.com/HL7/fhir-reasoning-omop-ri. Last accessed January 20, 2023.

Such an ecosystem also allows for new technologies in privacy preservation to be employed, which we expand upon in the response to Question 6.

3. **SMART on FHIR APIs**: We seek input on how U.S. Government and external stakeholders might leverage SMART on FHIR APIs, and in what areas further extensions might be needed.

MITRE recommends that SMART on FHIR applications be used considerately. In situations where the application presents a departure from a clinician's usual workflow, capabilities such as FHIR Questionnaires may allow for similar functionality while allowing the EMR vendor community the latitude to present the query unobtrusively. MITRE has had limited success with SMART on FHIR back office or server applications and strongly endorses further research into the possibility of using such an approach to facilitate the generation of a common data set such as one conformant to OMOP CDM. Finally, MITRE strongly supports the exploration of SMART on FHIR applications to allow patients to participate in the clinical research network.

The creation of provider facing SMART on FHIR applications introduces the possibility to remove them from their usual clinical workflows, introducing additional barriers and burdens to providers. Early in the journey of developing mCODE, the MITRE mCODE team developed a clinician facing SMART on FHIR application that allowed a clinician to compare the patient they were seeing against patients like them. The application then showed the outcomes of those patients in a series of different treatment options, as well as the most common adverse effects with those options.[8] User feedback sessions suggested the usefulness of such data, but there was a desire to see the data in their routine clinical workflow instead of an extra application.

In the response to Question 2, we described the development of an mCODE Extraction application that leveraged APIs. That application is an example of a back office or server type of SMART on FHIR application, and we would expect the limitations described would apply to the deployment of any SMART on FHIR back office or server type application. We do however return to the prospect of OMOP-on-FHIR as mentioned in Question 3 and note that a SMART on FHIR application could prove to be an exciting way to support health systems in the development of data assets conformant to OMOP CDM without having to develop and maintain costly ETL processes. Given the tremendous potential merits of such an approach, MITRE encourages further research along these lines.

We also note the strong possibility of patient-facing SMART on FHIR applications for research purposes. MITRE has explored the development of applications that allow patients to review their own data, as well as in reviewing potential trial matches in the CodeX clinical trial matching use case. We explore this topic further in the response to Question 6.

4. **Clinical Decision Support (CDS) Hooks**: We seek comments on how the HL7 CDS Hooks specification might be used to support clinical research, for example by creating prompts within the practitioner workflow during interaction with patients; and any advances that might be needed to support the use case described above.

[8] FluxNotes. 2023. GitHub, https://github.com/FluxNotes/flux/releases. Last accessed January 20, 2023.

MITRE recommends the cautious use of CDS Hooks in interventional trials where site to site variation can be accounted for or otherwise eliminated from the study design.

CDS Hooks support intervening into a clinician's workflow, with support for varying options of triggers and interventions. In the context of trial matching, MITRE has considered the use of CDS Hooks to support patient enrollment. The CodeX clinical trial matching use case is still conducting early investigations into the use of mCODE in their use case and is not yet to the point of interacting directly with patients.

A CDS Hook might also be considered to support provider workflow during a trial. Such prompts should be carefully designed, ideally with the delineation of roles at participating trial sites in mind. For example, in our ICAREdata studies many sites employ clinical research nurses to answer or review the ICAREdata specific questions, while other sites directly incorporate those same questions into their provider workflows. Such site-specific adjustments were prohibitively costly in the early setup of ICAREdata. In part, this led to the partnership between MITRE and Epic to introduce a standard set of locally adjustable tooling to capture mCODE for ICAREdata.

Again, these considerations are in support of interventional trials. For observational research, the Argonauts accelerator's efforts in researching and developing subscription-based workflows might be a more natural place for pilot exploration.[9]

5. **Operationalizing protocols of varying complexity**. As noted above, emergency clinical trial designs could range from relatively simple protocols to more complex studies involving the evaluation of investigational agents.

MITRE recognizes that the technological capabilities that could enhance observational and interventional trial designs have overlaps but are not in full agreement. Given the greater opportunity for standards and interoperability to facilitate observational trials, as well as the importance of observational trials in emergency clinical scenarios, we have chosen to prioritize technologies in this response that support such efforts.

In the responses above, MITRE has delineated between observational and interventional trial designs. We do so, recognizing that there is overlap but not complete agreement between the needs of observational research efforts and interventional ones. As noted in the response to Question 1, the presence of a common data model and research network led the OHDSI community to answer several important questions expeditiously and early in the pandemic. Topics we consider later in this response, such as privacy preservation and statistical methods for deidentification, are also more likely useful in observational research due to the greater possibility for such studies to be conducted using deidentified data.

Interventional trials may benefit more from technologies designed to intervene upon the provider workflow such as CDS Hooks or provider facing SMART on FHIR applications. Such tools may

[9] Clinical Data Subscriptions. 2023. GitHub. https://github.com/argonautproject/subscriptions. Last accessed January 20, 2023.

facilitate data capture, but as noted in the responses to Questions 3 and 4, our experience shows greater end-user satisfaction and adoption of interventions when the tools can be adapted to local variations in clinical workflow.

While both observational and interventional trial designs are critical to the development of medical knowledge, the majority of trials MITRE supported throughout the pandemic caused by SARS-CoV2 were observational. We therefore conclude that while both types of trials would benefit from further piloting and advancement, when considering emergency clinical trials piloting capabilities that primarily benefit observational research are of greater utility.

6. **Consent, deidentification, return of results**. The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution.

MITRE recommends prototyping and piloting of a clinical research network that takes advantage of recent advances in privacy preservation, data obfuscation, and observational research methods. MITRE further recommends that such a network permit patients to access their own record and to opt into or out of data sharing as they wish. MITRE has also conducted research using FHIR Consent resources, and this is an active area of investigation at this time.

The MITRE Corporation has explored several capabilities to facilitate the collection of consent and authorization. In our research on patient data management, we developed a patient data use agreement available both in full and an abbreviated, graphical view targeting a fifth grade reading level.[10] We have also explored the collection of consent directly from patients using web and mobile technologies similar to those used by the Sara Alert™ application for secure monitoring and reporting for public health.[11] This framework served over 8 million persons in U.S. territories during the COVID-19 pandemic and is openly and freely available.[12] Based on these experiences, MITRE concludes that technological approaches to obtaining patient consent can be performed at scale in support of the operations of a clinical research network.

As mentioned previously, a patient-facing SMART on FHIR application is of particular interest in such workflows. Such an application could display both textual and graphical versions of consent forms such as the ones demonstrated in the above patient data manager examples. These capabilities will allow under resourced entities to make more informed choices regarding their data and participation in clinical research.

MITRE has conducted several experiments on the subject of deidentification of electronic records or privacy preservation:

- MITRE developed the MITRE Identification Scrubber Toolkit, an open source and freely available resource for identifying and redacting personally identifiable information.[13]

[10] Patient Data Use Agreement for the Patient Data Manager. 2023. GitHub, https://github.com/patient-data-manager/pdua. Last accessed January 20, 2023.

[11] Secure monitoring and reporting for public health. 2023. Sara Alert, https://saraalert.org/. Last accessed January 20, 2023.

[12] Sara Alert. 2023. GitHub, https://github.com/SaraAlert. Last accessed January 20, 2023.

[13] The Identification Scrubber Toolkit. 2023. SourceForge, https://mist-deid.sourceforge.net/. Last accessed January 20, 2023.

- MITRE has created the Synthea synthetic patient population generator.[14] Tools such as Synthea can be used to test clinical research trial designs for validity prior to exposing them to patient data.
- MITRE has developed an internal prototype of a privacy preserving clinical research network using an open-source capability.[15] In such networks, data remains secure behind firewalls at originating locations. Each location configures their firewall to permit or reject queries based on that organization's legal environment and risk management procedures. Our experiments replicate successes in Europe, demonstrating that such networks are viable for preventing the unintended transmission of protected information.
- MITRE has conducted research into methods of Privacy Preserving Record Linkage (PPRL). In PPRL, data partners use an encryption key provided by a key escrow to obfuscate personally identifiable information. The obfuscated data is then linked by a third-party linkage agent, who creates a unique identifier that can be used to link data across the network.
- MITRE has also conducted independent research into newer statistical methods for privacy preservation in observational research, some of which are implemented by networks such as those enabled by federated learning networks or in the tools and libraries available via the OHDSI community.

As a result of these investigations, MITRE concludes it is possible to create a research network where privacy preservation occurs as a function of the network. Such a network would have several potential advantages:

- Under-resourced health care settings that cannot feasibility obtain expert reviews of incoming data requests could participate in select research efforts.
- Any agency, researcher, or citizen scientist with access to the network and a trial's code could replicate the experiment independently.
- In emergency clinical situations, it would be possible to develop ongoing and continuous monitoring of key parameters. It would also be possible to rapidly perform A/B type experiments and otherwise gain critical knowledge for early treatment.
- The network could be programmed to only permit queries that adhere to a given set of regulatory and ethical guidelines, either as a function of the entire network or under local control.
- Penetration testing, conformance testing, and other quality and security enhancing measures could be performed uniformly across network participants.

MITRE notes that several of the pilots and prototypes investigated toward deidentification of electronic records or privacy preservation function on data at rest (e.g., in relational data stores or free text files). While these capabilities themselves may not be directly applicable to data assets made available via FHIR resources, they serve as exemplars of how a network could be constructed using such principles. In the case of data conformant to the OMOP CDM, this is routinely stored in a relational data structure and would be more amenable to the aforementioned capabilities.

[14] Synthetic patient and population health data for the state of Massachusetts. 2023. MITRE, https://synthea.mitre.org/. Last accessed January 20, 2023.

[15] DataShield. 2023. DataShield, https://www.datashield.org/. Last accessed January 20, 2023.

OFFICE OF SCIENCE AND TECHNOLOGY POLICY (OSTP)

Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Introduction

The need for agility and speed have never been more critical to develop effective treatments, inform business strategies, and enable companies to reduce risk associated with future global events. In addition, market forces continue to drive the need for rigorous cost controls and strengthening business resiliency, while innovation demands new levels of flexibility in the core of clinical research. Key to future success is adopting a flexible and secure IT architecture that supports open collaboration models, streamlines the availability of data and facilitates advanced analytics at scale, and allows for the integration of healthcare-related solutions. The long-term success of digital transformation requires the industry be willing to reinvent business models, modernize and streamline operations, improve patient/provider engagement, and ultimately reduce time-to-market for life-saving treatments. Merative Clinical Development, previously part of IBM Watson Health, is a solution that has spent over ten years in the market offering a forward-thinking solution that operates in a different, more efficient and ecosystem-enabled way. The solution is a unified, cloud-based clinical data management and acquisition platform with customizable modules that can be tailored to the unique needs of clinical trials. With Merative Clinical Development, sponsors and trial owners are in full command of every aspect of their clinical trials and research – from designing workflows and forecasting costs, to building diaries for study participants. Merative Clinical Development empowers users to take control in every stage and the solution is designed to help accelerate trial outcomes with confidence.

The multi-step use case

1. A U.S.-level governing entity would oversee development of a clinical trial protocol for broad distribution across clinical trial networks and sites.

    Merative Response: This requirement generally aligns with the current operational needs for executing global research. However, there is a unique need in that most research conducted by a sponsor is executed under the technology architecture of the lead/one sponsor and made accessible to a network of providers/sites who are conducting the study visits and internal/outsourced organizations who execute the trial. As laid out here, the need to be able to access providers/sites and health networks regardless of sponsor-affiliation, and even to have multiple sponsors able to engage in a single "trial" system, would be critical. It will be necessary to have a solution that allows for any number of organizations to be able to access and conduct research within a single trial, regardless of affiliation. This exists currently with the Merative Clinical Development solution, where the platform can be made available to clinical trial network and sites with a userid/password and web enabled browser. The platform operates on a single code base across the entire unified platform and includes the

full integration of all its modules including EDC, randomization, inventory/dispensing management, medical coding, eCOA, eConsent, and more.

2. Study sites would enroll participants in the trial (potentially using software mechanisms that can alert sites to potential subjects for a specific protocol in a manner that increases the diversity of trial populations). Sites would obtain appropriate e-consents and authorizations from participants.

Merative Response: This requirement is aligned with currently available solutions for executing global research. The recommended approach would include e-Consent. For example, Merative's eConsent module allows for the enrolling of trial participants to be completed in only a few clicks and the e-consent module enables quick/easy authorization from participants without requiring an in-person visit. Different consent forms can easily be rolled out to meet the needs of each site, state, region, language, or country as applicable. There is also flexible customization that can include alerts via email and mobile phone.

3. Clinical trial data is typically sent to the trial sponsor though an electronic case report form (eCRF), which is the record of data that is required under the protocol to be captured for each trial participant. A data element in an eCRF is the smallest unit of observation for a particular subject.

Merative Response: This requirement is aligned with currently available solutions for executing global research. The real industry challenge lies in the often-lengthy time and effort it takes to build and deploy a study, with a major hurdle being downtime related to deploying ongoing updates as study changes are required during execution. It will be important to work with solutions that allow for quick and first-time right deployments. For example, the Merative eCRF design tool enables a study to get up and running in as quick as 4 days. Once running, changes to the design of an eCRF can be made in as little as 30 minutes without significant disruption to the trial. This is important in an emergency use due to the constantly changing information.

4. The eCRFs would be transmitted electronically via common APIs to the sponsor.

Merative Response: Open APIs are critical to any solution to enable data to flow on a regular, near real-time basis. In an emergency situation, delays in data receipt and review can be detrimental to the well-being of a clinical trial. Review of data cannot wait on development of one-off data transfers or trial based APIs. Solutions utilized in these types of trials should allow for the flow of data from start to end without delay. As an example, Merative Clinical Development offers a standardized ODM-based API available for access on day 1 of a trial, but also offers the ability to generate on-demand or scheduled file-based data transfers so that even those parties that are unable to call an API are able to review the data.

5. The study site's health IT system would present the eCRF content to clinicians in a manner that expedites data collection and (ideally) fits within clinician workflows.

Merative Response: The current way to solve for this is to provide access to the eCRF via a web browser and an internet connection, the way this is done with Merative and other systems. However, this does not meet the requirement to flow the eCRF data through the EHR/EMR system in order to make things accessible in clinician workflows. There are systems, such as Merative, that can enable this type of bi-directional flow via data integrations with ecosystem providers.

6. As the clinician obtains data elements to complete the eCRF, that data would be captured in the patient's electronic health record.

Merative Response: There are some industry examples of EHR/EMR<->EDC flow. For Merative, this can and is currently enabled via data integrations with ecosystem providers.

7. The clinical trial data would also be sent to a central data repository or small set of data repositories for researchers to analyze. It would be sent via common APIs so that researchers can easily interpret the eCRF data elements. Commercial cloud solutions are likely to house the data repository or repositories. Nonetheless, we would like a solution that would work across multiple cloud vendors.

Merative Response: The unique item in this requirement is the need to make data accessible from the central repository to a variety of researchers. With most industry solutions, data has to be copied/moved from a sponsor's controlled data repository to a shared data repository to support multi-sponsor access. With Merative, data can flow via Open APIs to a central repository or multiple as needed and be made accessible to any authorized user. In addition, within the Merative platform, access to the ongoing trial data can be made available to clinical trial network and sites with a userid/password and web enabled browser regardless of sponsor-affiliation.

8. For the purposes of this RFI, we are interested in the feasibility of all steps in the above hypothetical use case; we would also like input on how much of the use case could be operationalized in a pilot or demonstration project that might move forward in a timeframe of 6–12 months from the close of comments on this RFI.

Merative Response: As addressed in the individual answers, many of these requirements are currently standard in the industry, however the need to provide cross-sponsor, provider, and researcher access to the trial and resulting data is a unique need. Merative Clinical Development is unique in that the solution is already enabled to meet most of these requirements today.

The value of designing a pilot or demonstration project to operationalize data capture in the near term

1. United States Core Data for Interoperability (USCDI). We seek input on how U.S. Government and external stakeholders might leverage USCDI and future extensions of USCDI standards (such as USCDI+, an extension that supports federal partner program-specific requirements) to support emergency clinical trial research. It would also be helpful to receive comment on areas in which additional extensions might be necessary.

Merative Response: We do not have experience working with USCDI standards, however assuming the use case of USCDI (electronic health record data standardization) would remain distinct from that of CDISC (the regulatory submission standard for clinical trials), a mapping extension would be beneficial to ensure data collected in the USCDI standard could be programmatically mapped to the required submission standards.

2. HL7 FHIR APIs. We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-

emergency phase, and in what areas further advances might be needed. Specific topics in this connection include:

a. Use of an API that supports FHIR Bulk Data Access to support clinical research; whether bulk data exports from EHR systems can be used to support certain clinical trial protocols.

Merative Response: This is not currently the industry standard in terms of data source for EDC, however, there are proven case studies in the industry of successful data flow EHR/EMR<->EDC. Merative currently supports this use case via use of the data migrator that can ingest FHIR JSON. There are two challenges with flowing data into EDC directly from health records. First is that health record data is quite "dirty" and often clinical results data is found in unstructured clinician notes. The second challenge is that there is a need to control PHI/PII that may be housed in the health records to prevent that data from surfacing in the EDC. For these reasons the current industry approach is to ensure there is a mapping intermediary control in the data flow between the EHR/EMR and EDC systems.

b. Use of the FHIR Questionnaire and QuestionnaireResponse resources to support clinical research.

Merative Response: The FHIR Questionnaire is a structured approach to the metadata that defines how the content is presented to a subject. Within clinical trials today, validated instruments are already widely utilized to collect quality of life as well as specialized responses – and are based on medical research. A modification of these instruments would require an industry shift as alterations to the design are typically not allowed per licensing agreements. A marrying of the two would be required to ensure ease of mapping or each questionnaire would need to be created from scratch, further delaying the start up of an emergency use clinical trial.

3. SMART on FHIR APIs: We seek input on how U.S. Government and external stakeholders might leverage SMART on FHIR APIs, and in what areas further extensions might be needed. It would be helpful to receive comments on:

a. The most promising ways to create SMART on FHIR technologies that are portable across different institutions and EHR systems, but also provide adequate functionality to support emergency clinical trial research.

b. Whether the portability of SMART on FHIR tools provides a way to reach institutions and sites that have limited information technology resources; any promising ways to use SMART on FHIR to expand clinical research into underserved settings.

Merative Response: For healthcare systems that are underserved or work with underserved populations, the ability to afford and enable complex and often expensive EHR/EMR systems and patient applications may be the largest hurdle. Many patients may not have access to the internet nor have a location where they can easily have healthcare related conversations, while many may not seek healthcare until they are in an acute care scenario. Underserved populations can be reliant on emergency rooms for what is often considered

primary care related healthcare. And for providers who may still leverage paper records or patients who may not have access to nor understanding of their health records, the solution would likely be better focused on using the current technology to create simple and inexpensive cloud-based solutions that allow underserved providers and populations to participate in a 100% BYOD model, whether that be the emergency room, their local pharmacy or their own phone/tablet. This requires a significantly different approach to collecting, storing and accessing the data. The existing data standards are likely sufficient to enable this type of solution today with the intent to evolve the standards capabilities as the shift to more direct access becomes normalized.

4. Clinical Decision Support (CDS) Hooks: We seek comments on how the HL7 CDS Hooks specification might be used to support clinical research, for example by creating prompts within the practitioner workflow during interaction with patients; and any advances that might be needed to support the use case described above.

> Merative Response: There is an opportunity to alert clinicians of active clinical trials if the CDS Hooks were to be linked with https://clinicaltrials.gov/ (for example). Particular keywords – COVID positive – could show a listing of available clinical trials in the area, prompting the clinical to review these to see if a patient could be eligible.

5. Operationalizing protocols of varying complexity. As noted above, emergency clinical trial designs could range from relatively simple protocols to more complex studies involving the evaluation of investigational agents. We would appreciate comments on the following topics:

a. Whether any of the tools described above might be particularly well suited for certain types of studies.

b. For example,

- i. Whether a bulk FHIR API export could be used to gather data for a simple trial protocol that is relatively close to the standard of care for a particular condition.
- ii. Whether a FHIR Questionnaire/ QuestionnaireResponse or a SMART on FHIR form would be useful in capturing data for a more complex protocol, such as one that involves an investigational agent.

c. Any technical limitations that we should be aware of regarding use of the above tools to operationalize clinical trial protocols.

> Merative Response: Comments have been included above for each tool as to the challenges of using these tools within the current clinical trial landscape.

6. Consent, deidentification, return of results. The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution.

a. In light of this, we seek comment on how the tools described above can be used to obtain, collect and/or manage any required informed consents and/or authorizations from patients or individuals in accordance with applicable regulations.

> Merative Response: There are many existing tools in the market that specialize in collection of consent for clinical trials in an electronic format meeting applicable

regulations, of which most (if not all) do not currently utilize the above tools. Should the industry work to create alignment, the electronic consent tools could be embedded within patient EHR portals that present potential clinical trials (as put forward by the clinician), thereby opening a larger population for recruitment.

b. We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information. It would be helpful to receive comments on which deidentification and protection approaches are sufficiently mature to support a pilot effort in the near term.
c. Ideally, patient authorization would allow clinical trial data to be used for additional research beyond the original study. We would appreciate input on how the content collected for consent and authorization as well as the interfaces with deidentification technologies should be designed to enable flexible and responsible reuse of clinical trial data.

Merative Response: Clinical trial consents would need to have standard language covering their clinical trial data being used for additional research, with the ability for the subject to opt out. Clinical trial data is deidentified by design (e.g., subject IDs instead of names, redacted source documents).

d. We seek comment on any technical capabilities that could support return of results to study sites or participants, where appropriate.
e. We seek comment on any regulatory or ethical guidelines that are relevant to patients' consents and authorizations under the use case described in this RFI, and on ways in which technical solutions might help ensure adherence to applicable regulatory or ethical guidelines.

7. User interface and experience. With all of the above technologies, we seek input on:
a. The best way to optimize the experience of health care providers, administrators, and other users, so as to maximize the utility and uptake of the product.
b. To the extent a particular form, app or other tool requires input from a health care provider or other user, the best ways to increase the likelihood that users will actually provide that input. It would be helpful to receive comments on methods that are available for completing empty fields after the fact, or otherwise managing any missing data.
c. For clinicians and health IT users: what existing tools, apps, or processes you have found most usable and why.

Merative Response: In response to each of the above, we have seen that adding new tools, logins, technologies, or equipment comes with a learning curve, so introducing new tools when attempting to quickly start up a trial could cause delays and additional frustrations of clinicians already under duress. A good approach is to provide user friendly solutions that can be accessed using the same device or browser they are already utilizing. Providing a unified platform approach such that a user is able to login, enter data, view report, randomize a subject, enter lab results, or activate a questionnaire reduce the user burden. This has been our approach at building out Merative Clinical Development to provide a single unified

solution accessible to users on the device they feel comfortable and eliminating the need to switch to something else based on an activity.

8. Capturing data elements required for clinical trial protocols.
    a. We seek comment on the most promising technical approaches that would leverage common APIs to translate a particular clinical trial's data elements into data elements captured by user-facing tools (e.g., FHIR Questionnaire feeding into a SMART on FHIR form or application).
    b. If a tool such as a FHIR Questionnaire, FHIR Questionnaire Response, or SMART form or app is used to capture required data elements in this way, we seek comment on whether that creates an effective method for ''pushing out'' a research protocol to investigators and sites.
    c. It would be helpful to receive comments on how best to ensure compliance with regulatory requirements for eCRFs when designing interfaces for data capture.

9. TEFCA and QHINs. As noted above, TEFCA is in the implementation phase at this time. In the future, the TEFCA QHINs are expected to support implementation of the FHIR APIs (see the ONC Recognized Coordinating Entity's January 2022 FHIR Roadmap for TEFCA Exchange6). We would appreciate comment on the opportunities and challenges regarding development of API implementations toward the use case described above, particularly given the current status of TEFCA and QHIN participation. Specific topics in this connection include the following:
    a. Certain policy and/or technical constraints will need to be specified for currently authorized Exchange Purposes under the Common Agreement (e.g., Public Health). We seek comment on which of these constraints will also be applicable to a future research-focused Exchange Purpose.
    b. Opportunities that may exist for using the initially authorized Exchange Purposes to accomplish the use case described in this RFI.
    c. How the Public Health Exchange Purpose could be used to advance the goals of this RFI; what aspects of the use case described above might fall within the scope of the Public Health Exchange Purpose.
    d. How a future research-focused Exchange Purpose could be structured to advance the goals of this RFI.
    e. Other opportunities or constraints related to TEFCA that should be considered with regard to this RFI.

10. Emerging technologies. We welcome comments on any future technological developments we should anticipate. Relevant technical developments include but are not limited to differential privacy; federated machine learning; other technologies referenced in the recent OSTP RFI related to privacy-enhancing technologies (PET) (see Federal Register: Request for Information on Advancing Privacy-Enhancing Technologies); and technologies outside of the PET space. Specific topics in this area include:
    a. How future technologies might affect the use case and underlying assumptions laid out in this RFI.
    b. How future technologies might change the nature of the software architecture, data architecture, or potential data collection solutions for clinical trials.

11. Pilot or demonstration project. We seek comment on how the U.S. Government can best work with external stakeholders and developers to develop a pilot or demonstration project that will operationalize clinical trial data capture and serve as a basis and model for data collection in the event of an emergency. This pilot or demonstration project could also potentially support clinical research in the pre-emergency phase. Specific topics include:

a. Whether data can be managed through a central repository or small set of central data repositories; options for cloud-based data storage.

Merative Response: Merative Clinical Development provides a cloud-based solution where all clinical trial data can be entered by clinicians (or patients) that can be made accessible to a variety of parties on demand.

b. Technical options that might hold promise in the short term to enable researchers from diverse locations to analyze the data collected from multiple clinical trial sites. We also seek comment on any additional options that should be considered in the long term.

Merative Response: Merative Clinical Development is a globally available solution that has been used on over 3600 trials to date and is accessible by only a web browser and internet connection. In addition, for data entry by sites, Open APIs and mapping tools are available to help facilitate data transfer from other solutions.

c. Whether any parts of the pilot would be appropriately supported as

- i. A demonstration project with commercial partnership.
- ii. A public-private partnership.
- iii. An agency-funded program.

12. Specific commercial capabilities.

Merative Response: In summary, Merative Clinical Development's cloud-based clinical data acquisition and management system is poised to participate in a demonstration project to illustrate expeditious study builds for emergency studies. Using Merative's APIs, integrations with other systems would provide the sharing of information needed to realize the potential for clinical trial data sharing for additional studies per the consent of the subjects involved.
The challenge is the universal interoperability of all EHRs and EDCs, as well as those sites that do not currently have an EHR.

Merative is open to a demonstration project with a commercial partnership for the purposes of setting the bar for emergency clinical trials, and working with connections to EHRs and participation in QHIN for the purposes of exchanging data in a secure way for the betterment of trial recruitment, participation, and future research.

Datacubed Health

Proposal for Study: Emergency Clinical Trials RFI – Part 2

Sponsor: U.S. Government Office of Science and Technology Policy (OSTP)

Date: January 26th, 2023

Brett Kleger
Chief Executive Officer
e: Brett.kleger@datacubed.com
p: (484) 633-1849

Version 1

EXECUTIVE SUMMARY

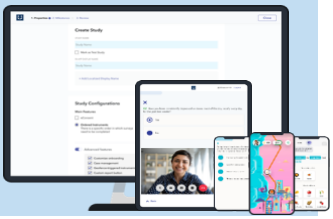
Datacubed Health ('Datacubed') is a pioneering patient engagement and data collection company designed for decentralized and hybrid clinical trials. Datacubed's mobile technology platform combines behavioral science with a SaaS technology that can be deployed in hours and is designed to optimize and simplify clinical trial participants experience and adherence.

Datacubed's mission is to ***Advance Health Access to Everyone, Everywhere***. This aligns directly with the goals of the RFI issued by the OSTP, specifically ...

- **Diversity, underserved communities** – the patient facing app is deployed via mobile devices. The trial is essentially brought to the patient, so they are not required to visit offices, take time off from work, or be otherwise inconvenienced. If a patient does not have a smart phone, Datacubed provides a device to the patient. The application also does not require a consistent Internet or Wifi access, as it may be used offline.
- **Emergency usage** – most technology providers in the industry require custom-coded solutions that take weeks or months to deploy. In contrast, Datacubed's solution is designed as a multi-tenant solution with an intuitive and flexible administrative interface that allows a study to be set up in minutes or hours, allowing for immediate deployment in the case of emergency usage/outbreaks.
- **Outbreak signals and indicators** – unique to Datacubed, the app includes geofencing capabilities to identify when individuals have entered a medical facility. This has been deployed on vaccine studies to signal when a patient enters a facility, followed by confirmation text with the patient. In an outbreak, this may be configured to signal when a patient enters a facility followed by confirmatory questions.
- **Large scale application** - Datacubed's solution can handle unlimited users and has proven scalability for 10s of thousands of patients, if not more. Combining a state-of-the-art technology infrastructure with the scale of Amazon Web Services (AWS) provides the scale and reliability required for widespread emergency usage.
- **Regulatory** – data collected by the Datacubed platform is regularly used in FDA or EMA submissions for clinical product approvals. Thus, Datacubed is well versed in regulatory requirements.

We welcome the opportunity to assist the OSTP and any partners in this important initiative and thank you for the opportunity to respond.

DATACUBED'S RECOMMENDATIONS FOR ENGAGING PATIENTS REMOTELY:



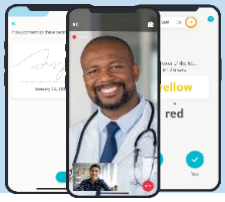
PATIENT REPORTED OUTCOMES (EPRO) / CLINICIAN REPORTED OUTCOME (CLINRO)

Patient access via mobile (iOS or Android) and/or web access; includes surveys (all question types), eDiaries, instruments, tasks, and caregiver access.

Site focused browser-based experience (tablet and/or web) for collection of patient data.

PATIENT ENGAGEMENT

Using a combination of motivational design and behavioral science, the platform features motivators, communications, and short-, medium- and long-term rewards that are intentionally designed to encourage adoption, engagement and use of our app.



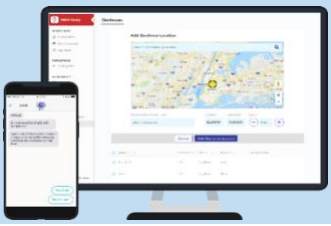
TELEHEALTH/ VIRTUAL VISIT

FDA-compliant telehealth feature available to the clinical and other study staff to videoconference with participants. Offers in-app scheduling and reminders for patients. Offers audit trails and HIPAA- and GDPR- compliant recordings for sites.

ECONSENT

This module allows delivery of predefined informed consent forms (ICFs), capture of signatures, and the delivery of comprehension questionnaires.

Consenting participants' signatures are fully FDA 21 CFR part 11 compliant.



GEOFENCING

Location services enabled geofencing around specific areas (e.g. hospitals) to detect health events and case alerts upon fence breaches.

INSIGHTS

Enable real-time reporting for active oversight on the progress of your studies. We have standard insights around instrument compliance, patient demographics and questionnaire detail, as well as custom insights around eligibility and scoring.

OSTP TOPICS

From the topics listed in the RFI by OSTP, Datacubed is responding to the following ...

> **7b. To the extent a particular form, app or other tool requires input from a health care provider or other user, the best ways to increase the likelihood that users will actually provide that input. It would be helpful to receive comments on methods that are available for completing empty fields after the fact, or otherwise managing any missing data.**

Datacubed Reponse: The platform consists of i) Linkt Admin (a web portal for project administrators), ii) Linkt Consent, and iii) Linkt App (the mobile app for participants and caregivers); all powered by a secure cloud infrastructure.

Within questionnaires, questions may be configured with a skip/branching or display logic. The survey/questionnaire builder currently allows for 14 different question types, ranging from a multiple-choice question type to a visual analogue scale (VAS) question type, and a free drawing question type, amongst many others. In a typical deployment, prebuilt and validated surveys can be selected from the extensive survey instrument library. Otherwise, the survey builder can be used to build precise copies of existing validated surveys – either by Datacubed Health staff or by the client. Datacubed Health also offers easily configured eDiary modules within Linkt. The configurable aspects of the module are the content, timing of deployment, and Gem reward.

Delivered by our breakthrough technology the Datacubed Health Patient Engagement & ePRO solution for clinical trials will create long-term engaging interactions with better questionnaire compliance, data quality, and reliable remote participation support and monitoring through:

- Expert guidance from service teams, trained to deliver a holistic solution for data collection powered by patient engagement and designed to deliver a seamless and enjoyable experience for patients, for sites, and for our sponsor partners
- Access to behavioral scientists and user experience designers focused on delivering seamless solutions for capturing high quality outcome data and improving overall trial adherence through engaging patient interaction with better patient experiences
- Flexible and versatile delivery tailored to the needs of your study protocol with increased efficiency for time to study start, reduced cost, and improved overall study team confidence powered by an industry leading configuration engine for study design and deployment
- A quality guarantee to meticulous delivery of the unique requirements per study design, mapping to data standards and clinical workflows, compliance to regulatory standards, and an ongoing focus on the individual critical success factors required for each unique protocol.

For additional sources:

* https://www.datacubed.com/scienceadvisors/

* https://www.datacubed.com/about/

Response to OSTP Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot

January 26, 2022

Re: Public Comment on Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot (87 FR 65259)

Submitted electronically to: datacollectionforclinicaltrials@ostp.eop.gov

Thank you for this opportunity to comment on the RFI regarding Data Collection for Emergency Clinical Trials and Interoperability Pilot. Epic is an electronic health records (EHR) developer based in Wisconsin. We provide the EHR platform for many of the most advanced healthcare research organizations in the United States and have extensive experience operationalizing many interoperability initiatives in complex environments. Our customer community includes 14 of the top 15 sites, by volume, for interventional clinical trials[1] as well as community health systems, FQHCs, retail pharmacies and other types of organizations that could contribute to better representation in clinical trials. In December 2022, organizations in the U.S. using our EHR platform, which includes research recruitment and study execution capabilities, were providing services for over 3.4 million active research patients.

We have extensive experience implementing FHIR, CCDA, and health industry standards required by ONC to support clinical interoperability. Over **3.74 billion patient records are exchanged per year** by providers using Epic, with 50% of exchange volume taking place with other vendors. We are active participants in the FHIR accelerator initiatives including Argonaut, CodeX, Da Vinci, Gravity, and Vulcan. We have worked with many external application developers as they develop FHIR-enabled apps to integrate into clinical workflow. Epic's comprehensive health record system can connect to any app that also supports FHIR to exchange health information, including but not limited to the U.S. Core Data for Interoperability (USCDI) data classes and elements.

Thank you for your commitment to addressing the need for a clinical trial deployment ecosystem better prepared to respond to a public health emergency. We enthusiastically support your effort to streamline the execution of clinical trials via the healthcare systems where patients are being seen. We encourage you to give more consideration to efficiently leveraging clinical care workflows while minimizing additional documentation burden on providers. Additionally, we encourage you to use more of the relevant FHIR resources, most of which are mandated under USCDI v2.

Please see our comments in the pages that follow.

Respectfully,

Nancy Smider, PhD
Director, Research Informatics
Epic

[1] Active interventional trials registered with ClinicalTrials.gov as of Jan 11, 2023

Design of the Protocol

As a general recommendation with respect to designing and developing a strategy for rapid clinical trial deployment leveraging health systems, any entity charged with this should include health IT (HIT) experts. In addition to long-standing experience incorporating externally developed content into clinical workflows in minimally disruptive ways, EHR vendors have extensive understanding of deploying health data interoperability at scale. Importantly, they have experience across many different types of organizations (e.g., community hospitals, independent practices, integrated delivery networks, retail clinics, safety net providers) in addition to the more traditional trial sites represented by academic medical centers.

Candidate Identification, Consenting, and Enrollment

Identifying candidates for enrollment

The use case outlined mentions that some EHRs may have mechanisms that can alert sites to potential subjects for a specific protocol to accelerate recruitment. Historically, a rate-limiting factor to rapid implementation of such features for a given trial is that the **inclusion and exclusion criteria are not specified in standardized healthcare interoperability terms and formats. Doing so would accelerate recruitment efforts.** For the trial protocol pilot, we strongly suggest conforming inclusion and exclusion criteria to USCDI data class/element representation to the extent possible. Subsequent screening can address any additional inclusion/exclusion criteria not covered by USCDI.

Consenting

The rapid execution of a trial would be facilitated further by using a central IRB, as contemplated in the related RFI (https://www.federalregister.gov/d/2022-23110), which could specify the *content* of a consent form, in multiple languages, and at an appropriate reading level for the general public. If an EHR natively supports consent collection, once the consent is collected at the site, the **FHIR Consent** resources should be used to communicate the authorization to the sponsor system.

You also ask about collecting consent in accordance with applicable regulations such that, ideally, the trial data could be used beyond the original study. Given the regulatory complexities around secondary data use for purposes beyond the specific trial, we recommend that this initial demonstration project stays focused and that the data use is restricted to the purpose of this project and the consent reflects that. Future projects can build on the technical framework established by the demonstration project and address consenting strategies that would permit use of data beyond the specific trial.

Enrollment

To track accruals across sites, the sponsor system or electronic data capture (EDC) system should also support the **FHIR ResearchStudy** and **FHIR ResearchSubject** resources. Some EHR systems can already provide patient-study status updates via FHIR (e.g., declined, enrolled, withdrawn, completed, etc.). These resources also facilitate tracking additional study-specific data (e.g., in combination with the **FHIR AdverseEvent** resource) so that adverse events, for example,

can be attributed to a particular study. Leveraging these FHIR resources lays the foundation to incorporate additional research-specific FHIR resources in the future. Additionally, they make it possible to support multiple studies over time that may involve the same patients.

eCRF Data Collection and Submission

Data collection and submission

As proposed in the use case currently, the eCRF seems to be the focus of the conceptualized workflow – i.e., how do we get a clinician to go to this eCRF form to fill in the trial data, without enough consideration given to how to minimize what is being asked of the clinician. We strongly encourage you to approach the design from a clinician-centric perspective. During a clinical encounter, many data points will already be documented natively in the EHR for that patient encounter (e.g., vitals, current mediations, demographics, etc.). A clinician should only be asked to enter data on an eCRF that cannot be prepopulated or retrieved via FHIR resources. This will reduce provider burden by asking them only to enter information not already captured in a patient's visit.

The RFI also asks specifically for feedback on the use of the FHIR Questionnaire resource(s) to support this trial use case. This seems to imply that you are considering using that resource for all of the trial-related data. We strongly recommend that you leverage FHIR resources appropriate to each data class to eliminate double documentation to the extent possible. The FHIR Questionnaire resource should only be used for data that aren't otherwise represented by separate FHIR resources. In "Input on Specific Topics" (below) we suggest several specific FHIR resources to consider, mapped against USCDI classes.

Privacy and de-duplication

Any de-duplication approach will require identifiable demographics at some point in the process. Local tokenization strategies such as Privacy Preserving Record Linkages (PPRL), though they seem "safe", may have additional unintended consequences particularly with respect to populations already typically underrepresented in clinical trials. Specifically, PPRL relies on a hashing mechanism that generates a token based on the patient's real name and DOB (or other identifiable demographics). This can disproportionately reduce matching success in some subpopulations such as the homeless (lack of address), certain ethnic groups (name structures), and women (last name changes) where accuracy and consistency of demographics is frequently more challenging. To combat the lower accuracy, several versions of the token may be created for combinations of Name/Age or Age/Sex to improve matching but simultaneously make it easier to triangulate the original patient identity. This is a challenge, and we encourage you to approach it with the recognition that de-duplication methods using demographics, even if obfuscated, will always leave an attack vector for re-identification. It is incumbent on a data aggregator to build multiple layers of safeguards to mitigate the risk. This includes keeping identifiable data and hashed tokens used for de-deduplication purposes technically isolated from the research data set, avoiding free-text data where identifiers are much harder to reliably identify and redact, and establishing a contractual/legal framework for the data consumers regarding legal and ethical use of the data.

Input on Specific Topics

How might USCDI be leveraged? Additional extensions/resources to consider?

UCSDI v2 data classes and the related FHIR resources that would likely be relevant:

USCDI Class	FHIR Resource
Patient Demographics	Patient (http://hl7.org/fhir/patient.html)
Vitals	Observation (http://hl7.org/fhir/observation.html)
Lab Tests/Results	Observation (http://hl7.org/fhir/observation.html); DiagnosticReport (https://hl7.org/fhir/diagnosticreport.html)
Medications for a patient	MedicationRequest (https://hl7.org/fhir/medicationrequest.html)
Condition	Condition (http://hl7.org/fhir/condition.html)
Procedures	Procedure (https://hl7.org/fhir/procedure.html)

Other key **FHIR Resources** (not currently part of USCDI):

- Consent (http://hl7.org/fhir/consent.html; date obtained, status, etc.)
- ResearchStudy (https://hl7.org/fhir/researchstudy.html)
- ResearchSubject (http://hl7.org/fhir/researchsubject.html)
- AdverseEvent (http://hl7.org/fhir/adverseevent.html)

How might SMART on FHIR be leveraged?

A SMART on FHIR app that leverages FHIR APIs to prepopulate fields will likely make it possible for a greater number of different types of institutions to participate in this sort of a trial more easily, though organizational governance will necessarily still play a role in enabling an external app to connect to their EHR. Neither SMART on FHIR, nor FHIR more generally, will eliminate the need for site IT effort completely (e.g., validation of mappings). Additionally, the entity developing the SMART on FHIR app will need to account for nuances among current implementations of FHIR (e.g., across major EHR vendors already supporting interoperability requirements). The potential viability of such a SMART on FHIR app could be demonstrated via a Connectathon, for example, as a complement to the proposed demonstration project.

How might CDS Hooks be leveraged? Could it create prompts for a practitioner?

A more specific problem statement would be very useful in order to provide a more prescriptive response for whether there is a role for CDS Hooks. If for recruitment, that may not be the preferred approach if a) it is disruptive to clinicians, and/or b) it isn't tightly integrated with whatever research recruitment and tracking capabilities the EHR already has natively. If it is to "remind" a clinician to complete something, there may be some value in that as long as it doesn't become unnecessarily disruptive, pull a clinician out of their workflow, or cause duplicative documentation.

Is a bulk FHIR API useful in this use case? Role for FHIR Questionnaire resources for more complex trials that involve investigational agents?

Given the transactional nature of the use case presented (requires consenting and potentially additional data collection at the point of care), bulk FHIR is unnecessary and we do not see any particular advantage to using it for the pilot use case proposed. Data acquisition via FHIR can be triggered efficiently during the clinical workflow itself. Bulk FHIR might be of use for additional data

retrieval at some subsequent timepoint for any patient who had participated in the trial, whether or not they were currently receiving additional care. That is not what the current use case proposes, but worth considering for subsequent phases of this effort.

As noted above, FHIR resources already supported by certified EHRs and fit for purpose to the data classes of interest for the trial should be used. The FHIR Questionnaire resources should be reserved for data that isn't otherwise represented by native FHIR resources. This should be the case for trials involving investigational agents as well. The eCRF should be configured to support these APIs. It then becomes a relatively straightforward matter for the site to implement the form (after confirming the mapping). The current proposed use case, however, does not mention whether an intervention is involved. If it is intended to support the needs of an interventional trial (e.g., vaccine or therapeutic interventional trial) the design should explicitly include this so appropriate workflow can be considered and relevant interoperable representations can be applied.

Capabilities that could support return of results to study sites or participants

To return overall study results to patients (not patient-specific test results), consider leveraging the **FHIR DocumentReference** resource. That can file back to the participant's record at the site. If this is instead referencing a specific participant test result, that result (if ordered in the EHR) would likely already be available in the EHR and not need to be "returned" to the site. If the intent would be to package up all of a patient's own results related to the study, that could be done and then returned via a PDF document to the patient chart, again using the **FHIR DocumentReference** resource.

Regulatory/ethical guidelines relative to consents

Assuming that consenting is required for the proposed trial, as we noted earlier the rapid execution of the trial would be facilitated by a central IRB which specifies the content of a consent form, in multiple languages, and at an appropriate reading level for the general public. If an EHR system already supports electronic consenting that is compliant from a regulatory perspective, we would urge you to encourage sites to leverage the EHR's e-consenting capabilities that their providers and patients are already familiar with. This would provide better workflow consistency for the site and for the patient, including compliant storage of the signed consent where it could be accessed by a patient in the same way they access other consent forms they've signed through the health system. The health system would also retain record of that signed consent form as well. Allow sites with an existing e-consenting platform to use that platform as long as it supports the **FHIR Consent** resources. This would ensure that information about the signed consent could be transmitted to the sponsoring entity for their records. The lead entity could provide an optional technical strategy for consenting to accommodate sites that don't already have e-consenting technology that would meet the interoperability requirements.

User interface and experience

We re-emphasize the importance of maximizing the use of data already captured in an encounter and minimizing additional data entry by clinicians. We strongly recommend against using hard stops within any additional data capture forms. In most cases, if form design is good, with only minimum necessary additional entry requested, and if reasonably embedded in clinical workflow, providers are likely to complete the necessary information. This is especially true if relevance to the public health emergency trial is clear. Interfering with the ability of a clinician to navigate as they deem necessary to care for a patient during that encounter because of hard stop in a data entry field is strongly discouraged except in cases where patient safety is a stake. Hard stops research data collection forms may contribute to clinician frustration and, potentially, lack of form completion at all. Warnings or reminders can be helpful but should be used with discretion.

With respect to the user interface and workflow design, recommendations and guidance are helpful, but the protocol should not dictate the specific user interface and workflow design in any system. EHRs, and even different organizations using the same EHR, may have specific strategies for workflow design and incorporating externally provided applications/forms.

Capturing data elements

Above we addressed how to better leverage already existing USCDI and FHIR resources. Regarding form deployment itself, if a form is provided via a SMART on FHIR app, the health system will need to validate the fields being used to populate the form as they deploy that the app. REDCap CDIS[2] provides an informative example of how this can be executed at the site level.

eCRF regulatory compliance requirements

Compliance to regulatory requirements should not be confused with user interface (UI) design and should not dictate the UI of the workflow. The protocol should define the specific regulatory compliance requirements that must be met and could provide suggestions for embedding within workflow. Allow UIs to be optimized by the sites who will be embedding the study. Different EHRs have different UI capabilities and flexibility. Attempting to force everyone, regardless of EHR platform, to execute the workflow the same way is unlikely to maximize adoption or usability. The design of a supplemental data capture form provided by a SMART on FHIR app would, of course, be up to the app provider but they should not attempt to determine the most efficient UI to guide the user through the workflow surrounding completion of that form.

Regulatory compliance requirements (e.g., metadata about how/when the data were acquired) need to be supported by the application you propose to develop leveraging existing interoperability approaches. FHIR resources (e.g., the observation resource) typically represent metadata regarding the source data, such as status, performer, performerFunction, etc. From a 21 CFR Part 11 compliance perspective, in July 2018 the FDA released a Guidance for Industry on the Use of Electronic Health Record Data in Clinical Investigations. In this guidance the FDA reiterated that "Under the ONC Health IT Certification Program, certified EHR technology would be in compliance with applicable provisions under 45 CFR part 170. EHR technology with certified capabilities

[2] https://projectredcap.org/software/cdis/

generally has clear advantages, because many of the certification requirements are aimed toward ensuring interoperable data sharing and enabling processes to keep electronic data confidential and secure. In particular, all EHR technology certified under the ONC Health IT Certification Program is required to meet certain privacy and security protection requirements for an individual's health information (see 45 CFR 170.314(d)(1) through (8) and 45 CFR 170.315(d)(1) through (11)). FDA encourages the use of such certified EHR systems together with appropriate policies and procedures for their use."[3]

It may be of use to have a site activation checkpoint to confirm that the EHR from which data will be retrieved meets current ONC Health IT Certification requirements.

TEFCA

The Trusted Exchange Framework and Common Agreement (TEFCA) was established by the Office of the National Coordinator (ONC) as a result of the 21st Century Cures Act, which directed that the ONC develop or support a trusted exchange framework, including a common agreement among health information networks nationally. TEFCA is not intended to replace existing exchange networks, but to complement them. It aims to do so by establishing a universal floor of nationwide interoperability and to enable organizations not currently part of a broader exchange network to more effectively participate in health information exchange. TEFCA can inform the development of a strategy around differences in interoperability implementation that may exist today.

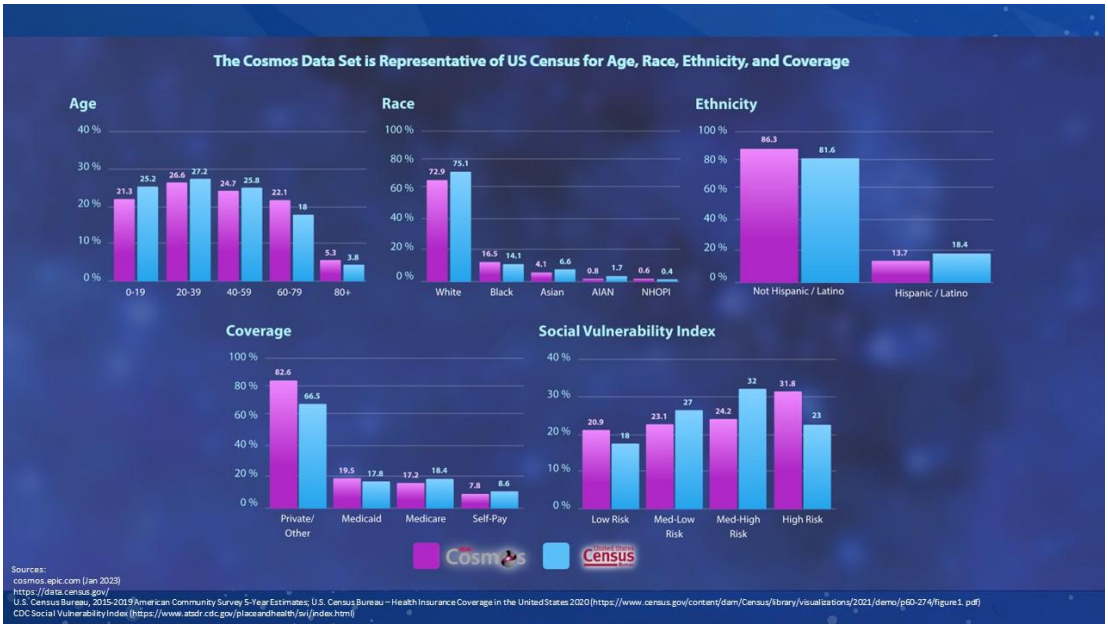
This RFI should be decoupled from input regarding TEFCA. While it shows significant potential, TEFCA is not yet live and its effectiveness remains to be established. TEFCA is designed to facilitate the exchange of data through QHINs, not the coordination of workflow or data collection necessary for a trial as outlined in this RFI. Additionally, as you note, exchange for the purposes of research is not a use case TEFCA is currently addressing (Treatment; Individual Access Services) nor one of the "next up" use cases (Payment, Health Care Operations, Public Health, and Government Benefits Determination)[4], though it has been noted as a potential for future consideration. One can imagine that TEFCA could evolve, at some point, to support data exchange for the purposes of research, though there are other regulations regarding use of health data for research purposes that add complexity and would need to be considered in conjunction. **This is a complex topic and deserves a dedicated RFI of its own, perhaps issued jointly with ONC.**

Specific commercial capabilities of note

For observational clinical data surveillance there are already existing real-world data resources that can be called into service. For example, as of January 2023, Epic's Cosmos (cosmos.epic.com) contains the deduplicated, de-identified longitudinal clinical data of approximately 176 million people in the United States.[5] The demographic profile of the Cosmos population closely mirrors the U.S. Census on a number of measures including age, gender, race/ethnicity, geographic location, and

[3] https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM501068.pdf
[4] https://rce.sequoiaproject.org/wp-content/uploads/2022/06/SOP-Exchange-Purposes.pdf
[5] https://cosmos.epic.com (January 11, 2023)

healthcare coverage. This relatively new data resource is already being leveraged by the CDC[6,7] and other stakeholders to rapidly answer research questions based large-scale national data.



[6] Plumb ID, Feldstein LR, Barkley E, et al. Effectiveness of COVID-19 mRNA Vaccination in Preventing COVID-19–Associated Hospitalization Among Adults with Previous SARS-CoV-2 Infection — United States, June 2021–February 2022. MMWR Morb Mortal Wkly Rep 2022;71:549-555. DOI: http://dx.doi.org/10.15585/mmwr.mm7115e2

[7] Shah MM, Joyce B, Plumb ID, et al. Paxlovid Associated with Decreased Hospitalization Rate Among Adults with COVID-19 — United States, April–September 2022. MMWR Morb Mortal Wkly Rep 2022;71:1531–1537. DOI: http://dx.doi.org/10.15585/mmwr.mm7148e2

These comments are filed by Stuart Buck and Betsy Ogburn (a bioscientist at Johns Hopkins) on behalf of the Good Science Project (a think tank) in response to two White House Requests for Information:

- Clinical Research Infrastructure and Emergency Clinical Trials, 87 Fed. Reg. 64821, and,
- Data Collection for Emergency Clinical Trials and Interoperability Pilot, 87 Fed. Reg. 65259.

As the White House has correctly noted, the current clinical trial infrastructure in the United States is "not well prepared" to perform large and rigorous clinical trials in the case of emergencies or pandemics. One particular instance of this lack of preparedness involves the design of trial protocols and the capture of data in an efficient and interoperable format.

At the Good Science Project, we believe that the White House has identified real problems. Nonetheless, the narrow focus on emergencies and pandemics is misguided. A robust clinical trial infrastructure (including data standards) cannot be a mostly-theoretical plan that is left on the shelf, to be used only in case of emergency or only for specific diseases. The only way that such infrastructure will be usable at all is if it is constructed and used *on an ongoing basis.* In that case, it could be readily redeployed in case of emergency; but if not otherwise in use, we would still be starting essentially from scratch in future emergencies.

I. The Goal Should Be To Create an Efficient Clinical Trial Infrastructure *Period*, Not Just For Emergencies

We should create an efficient clinical infrastructure to be used on an ongoing basis, not just in so-called "emergencies." This is true for multiple reasons.

For one thing, drawing a line between emergency and non-emergency settings is arbitrary. Are leading causes of death and morbidity–like cancer, heart disease, Alzheimers, and mental health disorders–less worthy of attention than pandemics? They may get less attention from journalists or politicians, but they remain leading causes of death and disability–year in, year out.

For another thing, drawing a line between emergency and non-emergency settings will be counterproductive and will only lead to failure. Clinical trial infrastructure is not like a hatchet kept behind glass, to be broken only in case of emergencies. Instead, clinical trial infrastructure

needs to be built out and deployed *to be sure that it works at all*, and then we need to *iterate and improve on an ongoing basis*. If left for so-called "emergencies," chances are good that any infrastructure or coordination protocols would fail, if only because no one had never worked out the kinks.

Thus, we recommend building systems for protocol harmonization, data sharing, and encouraging (or even engineering) large multi-site trials, and deploying them *now* for *all* important clinical questions. Not only will such efforts improve clinical research across the board, such an infrastructure would be ready to redeploy in case of a fast-paced emergency like a pandemic. Perhaps in "emergency" settings, one might want to force institutions and PIs to waive the right to first publication(s), but otherwise the infrastructure and data standards should be the same.

II. How To Improve Our Ability to Accumulate Evidence From Clinical Trials

A related problem with the current clinical trial infrastructure is that we are failing to fund high-quality trials that are coordinated in advance, as well as to properly pool and accumulate evidence.

First, we need to build infrastructure to tap into non-academic medical centers with clinical trial infrastructure. Pharma companies routinely contract with small medical centers around the country (including in rural areas) to run clinical trials. These centers have excellent trial infrastructure and, anecdotally, are often interested in academic research but have no entry point in most cases. NIH, perhaps through the CTSAs, and/or PCORI should develop a network of these sites that could be deployed to increase the sample size and generalizability of high-priority funded trials, and to have at the ready for "emergencies."

Second, ClinicalTrials.gov is not a sufficient data-sharing platform for federally funded research. Data that was paid for by the US government *must* be made maximally useful for improving public health and clinical practice. An idea from Barbara Bierer (and others) is that we should incentivize data "authorship" to give credit for producing data that is used for productive research down the road. For example, NIH, PCORI, and others could prioritize the creation of useful data over publication of first-author papers when evaluating PIs for future funding, because the production of data is what most advances our collective knowledge base.

Third, NIH could consider moving towards a contract, rather than grant, model of funding for some high-priority clinical questions. This is similar to PCORI's model, and would better enable NIH to make decisions on behalf of PIs, such as requiring multiple teams to come together in a multisite trial, enforcing the adoption of a common protocol, expanding a trial's footprint by enlisting non-academic sites to enroll more patients, etc.

Fourth, NIH should try to completely eliminate "small crappy trials." While there are many pilot and feasibility trials that may be worth funding, there are still too many "small crappy trials" that get funded for some reason. NIH should make a considered effort to fund fewer and better trials along the lines of ACTIV, CAST, WHI, NLST, and more. Single-arm trials (with no

randomization) should not be funded except in the most extenuating of circumstances. And in many cases, this may require lifting the NIH's de facto $500,000 limit on direct costs per year.

Finally, when it comes to data systems and standards, we should aim to 1) include at least the *possibility of randomization* at the point of care, so that many more doctors and health care facilities are enabled to participate in clinical trials with a lower barrier of entry; and even more ambitiously, 2) rebuild the electronic health care records systems in the US to tailor them for research rather than merely for reimbursement. The latter would make clinical trials far less expensive to run and would drastically lower barriers to pooling evidence across institutions. The Office of the National Coordinator (ONC) should make that their top aim and priority when it comes to interoperability.

castor

Castor response to

OTSP RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot

***Date**: Jan 26th 2022*

***Author**: Derk Arts, MD PhD, CEO*

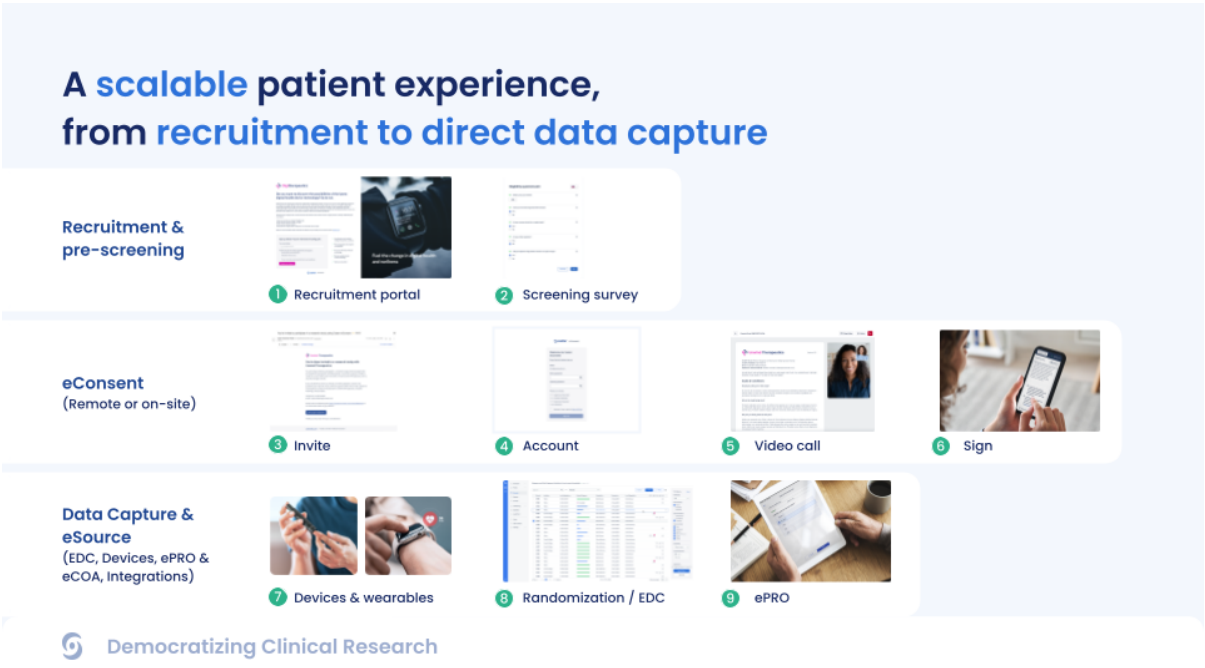
Introduction & Background

Castor is pleased to respond to the RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot. We understand the importance of enhancing U.S. capacity to carry out clinical trials in emergency situations and are committed to supporting the goals of the National Biodefense Strategy and American Pandemic Preparedness Plan.

Castor is a leading provider of clinical trial technology that automates the research process. Our platform is used by researchers and organizations all over the world to streamline and simplify the clinical research process. Castor's platform is designed to help researchers easily capture and manage clinical data. Our platform is user-friendly and intuitive, making it easy for researchers of all skill levels to use. We have a wide range of features, including electronic data capture (EDC), e-consent, ePRO, randomization and data management capabilities.

Our platform has already been used to power hundreds of COVID-19 trials globally, including the World Health Organization's (WHO) Solidarity trials. The Solidarity trials are Platform / Umbrella trials aiming to identify treatments and vaccines for COVID-19, with more than 35,000 patients randomized and 5,100 investigators at 550 hospitals across 30 countries

The experience we gained from working on the Solidarity trials gives us a unique perspective on what it takes to run a large-scale clinical trial in an emergency setting. We understand the importance of speed, efficiency, and data security when it comes to pandemic preparedness. Our platform supports HL7 FHIR standards and is compliant in any region in the world, with the exception of China. Our team is well-versed in the technical aspects of data collection and interoperability. With our experience and expertise, we believe we are well-positioned to help the United States build a clinical research infrastructure that is ready for any emergency.



Background on this RFP Response document

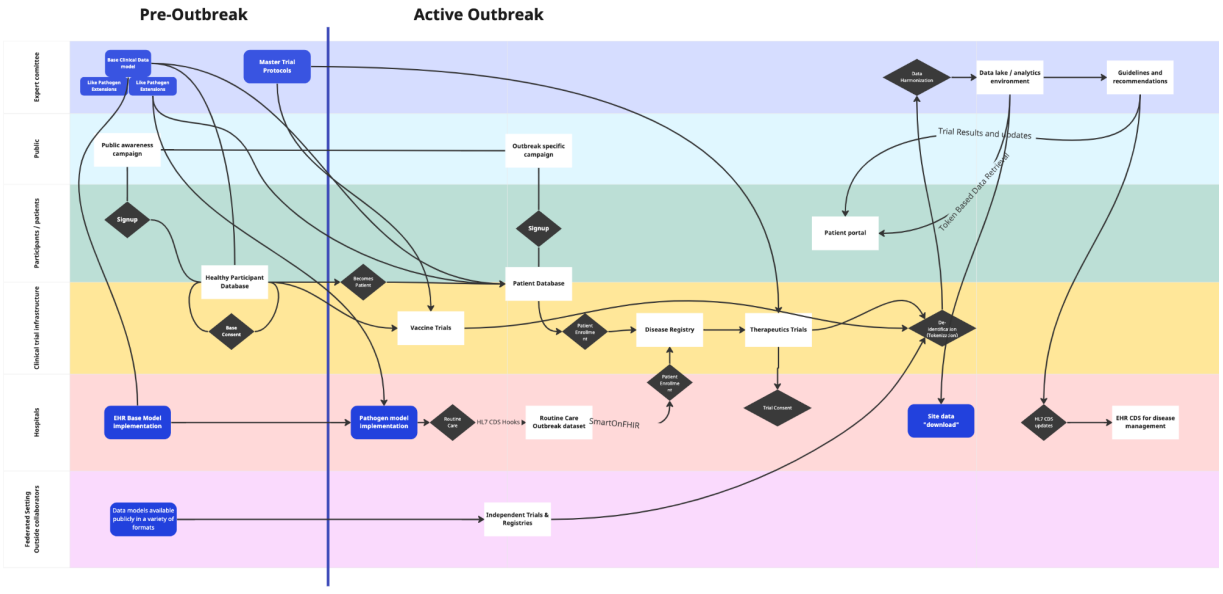
Castor understands the importance of simplicity and robustness in the chosen approach to ensure it can work well at a large scale and not fail due to over-engineered solutions. In an emergency setting, an anti-fragile approach is essential because everything that can break, will break.

We recommend a 3-tier approach to avoid depending on complex technology that is not fully implemented or tested in the real world:

- **Tier 1**: Direct to consumer model, with open Healthy Volunteer and Patient registries. This allows for quick and easy enrollment of participants and captures relevant data directly from the source.
- **Tier 2:** User-friendly clinical trial platform that clinicians can use for consenting and direct data capture of patient data. This ensures that the data being captured is accurate and relevant, while also reducing the burden on clinicians.
- **Tier 3:** FHIR-based architecture for capturing relevant data directly in EMR systems with minimal disruption. This allows for seamless integration with existing systems and eliminates the need for double data entry.

We will walk through the solution overview step by step to explain the suggested solution in more detail and how it addresses the goals of the RFI.

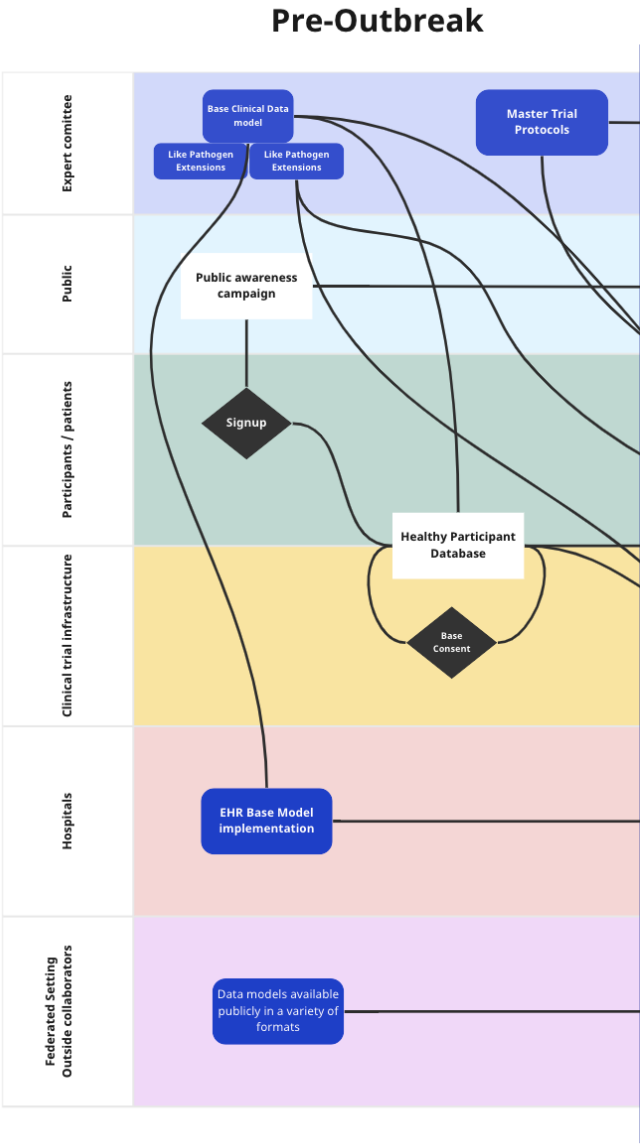
Solution overview



The (simplified) overview of the proposed solution.

Pre-Outbreak

Below we briefly discuss each element of the solution that we recommend would be prepared ahead of an outbreak



Annotated Data Models

Data models, developed by an expert committee, consisting of a Base Clinical Model that would cover any disease, with predeveloped extensions for the most likely pathogens. These would need to be annotated with common ontologies as well as CDISC CDASH concepts.

These models would ensure that the data captured is relevant and can be used to understand the progression of the disease in patients.

USCDI

In the context of emergency clinical trial research, USCDI and future extensions of USCDI standards, such as USCDI+, could be leveraged to support the efficient and accurate exchange of clinical trial data between trial sponsors, study sites, and researchers.

One potential area where additional extensions to USCDI might be necessary is in the area of patient consent and authorization. In emergency clinical trial research, it is important to ensure that patients are fully informed about the trial and that their consent is obtained in a timely manner.

Implementation and distribution of models:

- Implementation as a FHIR model in EMR systems: This will allow for seamless integration of the data models into existing EMR systems, ensuring that the data can be captured in real-time and used for research purposes.
- Implementation in a clinical trial platform such as Castor to support patient and clinician facing data capture: This will ensure that the data is captured in a consistent manner and can be easily shared between research teams.
- Publicly sharing the models in a wide variety of formats for broad support for EMR and EDC systems (e.g. CDISC ODM): This will ensure that the data models can be used by a wide variety of systems and can be easily shared between research teams.
- Publicly sharing metadata about running clinical trials using our FAIR Data Point infrastructure. With proper authorization, data could also be shared and queried in a semantically interoperable manner.

Master Protocol design

- Development of Master Trial Protocols for a potential Therapeutics and Vaccine trial: This will ensure that the research teams are ready to act quickly when a new pandemic occurs, and that the protocols are in place to ensure that the research is conducted in an ethical and efficient manner.

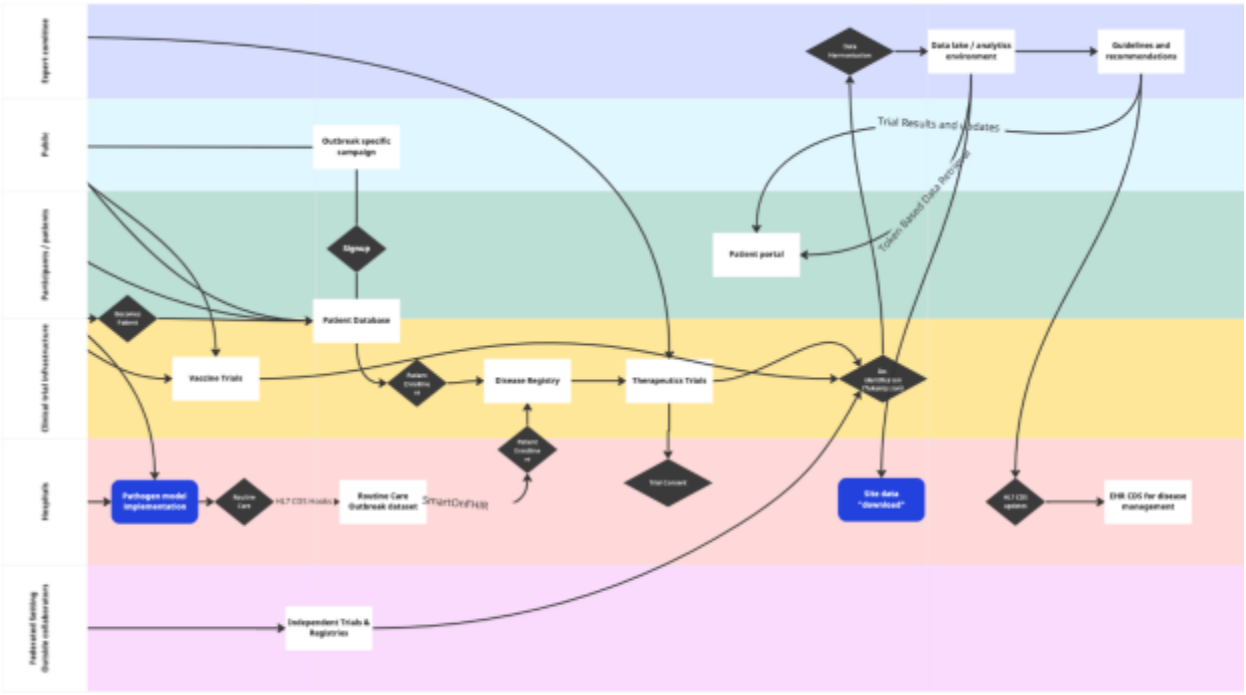
Healthy Participant enrollment portal

- A portal where volunteers can enroll to be contacted for a potential future Vaccine trial, or as a baseline dataset for when they get infected with the pathogen / disease: This will ensure that there is a pool of participants that can be called upon when a new pandemic occurs, saving time and resources in the recruitment process.

- Public awareness campaigns can be used to invite the public to enroll in the Pandemic Readiness Healthy Participant Portal: This will ensure that the portal is well-populated and that a diverse group of participants is enrolled.

It is important to note that these steps must be done in advance of a potential pandemic, so that the system is ready to respond quickly and effectively when the time comes. This will ensure that the data is captured in a consistent manner, that there are participants ready to enroll in a trial, and that the protocols are in place to ensure that the research is conducted in an ethical and efficient manner. In addition, it will also help to reduce the burden on the healthcare system by having a clear plan in place.

Active Outbreak



Identifying patients and enrollment into a disease registry
The goal of this section of the solutions is to have a robust, “anti-fragile” approach to building up a registry of patients and critical baseline disease data as rapidly as possible. The three Tiers below each contribute to this section.

- **Tier 1 - Patient Database:** In addition to the healthy volunteer database we recommend launching a patient database in which patients affected by the disease can enroll themselves. Castor successfully deployed such a portal for the WHO to register breakthrough vaccine cases.
    - 1.a Healthy volunteers that were previously enrolled can amend their record by recording they now contracted the disease.
    - 1.b The Healthy Volunteer and Patient Database can be used to identify and enroll into Tx and Vaccine Platform trials.

- **Tier 2 - Direct Data Capture Disease Registry:** Direct data capture by clinicians into a disease registry: to ensure we do not have to rely on the availability of an EMR system

and the implementation of data models in those EMRs, we recommend a globally available Direct Data Capture such as Castor EDC to offer any clinician, regardless of setting, a way of contributing to this project. This approach has been successfully used in both WHO Solidarity trials.

- **Tier 3:** The preferred long term solution that leverages a combination of pre-built EMR forms (based on the initially developed data models), SmartOnFHIR apps, and CDS Hooks to capture clinical disease data and alert physicians that their patient can be enrolled into the Disease Registry.

    - **SMART on FHIR APIs** could support this project by offering portability across different institutions and EHR systems. This would enable the sharing of patient data between different institutions and EHR systems, which is critical in emergency settings where time is of the essence. We believe however, that we cannot rely on this technology to be the only method of gathering data in this project as nation-wide EMR implementation may face a wide variety of technical and legal roadblocks.
    - **FHIR based CDS hooks** can be used to create prompts within the practitioner workflow during patient interaction, to remind practitioners to collect specific data elements required under the trial protocol.
    - **Through use of predefined data models** we can ensure compatibility of data captured in the EMR during the regular clinical workflow with the centrally maintained disease registry.
    - **Data Return** to patients can happen through their existing EMR patient portal, where the clinical trial platform pushes HL7 FHIR message back to the local EMRs.

Running clinical trials (e.g. Therapeutics or Vaccine trials)
The Castor platform has been used in over 14,000 trials worldwide. We propose to leverage the healthy volunteer & patient databases to enroll into a potential vaccine trial or disease registry respectively. This observational disease registry can in turn provide patients and baseline data to a Randomized Therapeutics Platform trial based on the prespecified Master Protocols.

Pilot Approach

Castor has successfully supported global Platform Trials in an emergency setting with the World Health Organization. We deployed the platform for them within 5 days and were randomizing patients within 2 weeks. This makes us confident we can deploy and dry-run this infrastructure within the 6 months time-frame.

- Castor can deploy the following capabilities out of the box, with minimal configuration
    - Enrollment portals for Healthy Volunteers and Patients
    - Observational Disease Registry
    - Clinical Trial Platform that supports Platform Trials (Vaccine & Therapeutics)

- Castor would partner for the following capabilities
    - Technology to connect to national FHIR APIs

We expect to be able to deploy a pilot infrastructure that can support the pre-emergency phase of this project, and act as a foundation for the full scale project.

Clinical Data would be managed through our CDMS platform, acting as a central repository. This solution would utilize cloud-based data storage as it provides the necessary scalability, security, and accessibility for large-scale data management.

Castor will provide a user-friendly interface that allows researchers to access and analyze the data. Our FAIR Data Point software, through which data and metadata can be made Findable, Accessible, Interoperable, and Reusable, can be leveraged to expose data and metadata of trials. This allows researchers to access and analyze data from various sites or trials in a secure, decentralized, and federated manner.

Partnership model

- Castor would be open to both a “demonstration project with commercial partnership” or as a partner in an “agency-funded program.”
    - We have successfully contributed to large funded projects in the past, for example in a “European Horizon 2020 SME Instruments Phase 2”

Suggested steps for pilot deployment

1. Develop a draft Emergency Protocol design, consider reusing SOLIDARITY or RECOVERY protocols.
2. Develop Basic Models for Pilot Project, consider using existing COVID models that have been developed.
3. Work with 2 or 3 mature Hospital Systems that have successfully implemented innovative solutions into their EMR systems. Rely on them to implement the models to capture data as part of the clinician workflow.
4. Castor will set-up the Healthy Volunteer Portal, Patient Portal, Disease Registry and a Mock Platform Trial.
5. By using COVID as a test Pathogen, this infrastructure can be tested in a real world setting.

Contact Details

- D.L. Arts, MD PhD, CEO
- derk@castoredc.com
- 201-2571783

Palantir

ORGANIZATION INFORMATION

Organization Name: Palantir Technologies Inc. (“Palantir”, “we”, or “our”)
Organization Type: Industry
Organization Description: Palantir is pleased to submit a response to the Office of Science and Technology Policy (OSTP) request for information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot. Palantir builds software that empowers organizations to effectively integrate data, decisions, and operations. Our clients include commercial organizations and Government Agencies with health and life science-focused missions. The Foundry configuration of the Palantir Platform (“Foundry”) has been deployed in emergency contexts to provide both operational and research capabilities. This includes during COVID-19, where Foundry was leveraged by HHS to monitor the spread of the disease, track the development and distribution of vaccines, therapeutics, and personal protective equipment (PPE). We built interoperability, privacy controls, data protection, and governance into our platforms since inception.

During the COVID-19 outbreak, the National Institutes of Health (NIH) turned to Foundry to quickly stand up the National COVID Cohort Collaborative (“NC3”), a secure platform through which the harmonized clinical data provided by contributing members is stored. The data itself can only be accessed through a secure cloud portal hosted by the National Center for Advancing Translational Sciences (NCATS). N3C is an open science community focused on analyzing patient-level data from many clinical centers to reveal patterns in COVID-19 patients. To create N3C, the community overcame technical, regulatory, policy, and governance barriers to sharing patient-level clinical data. In less than two months, N3C and Palantir configured solutions to acquire and harmonize data across organizations and created a secure data environment to enable transparent and reproducible collaborative research. Ultimately, N3C aims to help save lives by enabling collaboration among clinicians, researchers, and data scientists to identify treatments and specialized care needs and thereby reduce the immediate and long-term impacts of COVID-19.

As demonstrated by N3C and many other programs across the Federal Government, we strongly encourage the inclusion of commercial software in OSTP’s exploration of Data Collection for Emergency Clinical Trials and Interoperability Pilot.

RESPONSES TO QUESTIONS

1. United States Core Data Interoperability (USCDI)

Palantir encourages the evolution and adoption of standards and best practices that promote nationwide, interoperable health information exchange. These should expedite and improve the quality of clinical trials established in an emergency context. Beyond implementing recommendations on the nature of specific, necessary extensions (such as USCDI+), OSTP should ensure that any pilot capability regarding Data Collection for Emergency Clinical Trials and Interoperability demonstrate the ability to meet evolving interoperability standards and requirements swiftly and effectively.

For example, health IT solutions should demonstrate an ability to incorporate newer versions of Secretary-adopted standards and implementations as established in Office of the National Coordinator for Health Information Technology (ONC’s) Standards Version Advancement Process (SVAP).

While leveraging standards like USCDI serves as a foundation, these standards alone are insufficient to ensuring core data interoperability in current and future states. These standards are likely to evolve over time and so any system implementing them requires swiftly adaptable capabilities to support robust change management, otherwise the delivery of care risks falling behind the standards that aim to bolster patient outcomes. Adaptable system capabilities are further necessitated in the emergency clinical trial context, where rapid implementation capabilities for interoperability standards need to keep pace with improvements in scientific understanding of the pathogen, how it spreads, and the corresponding pathophysiology of infection.

2. HL7 FHIR APIs

In a research context, common data standards such as Fast Healthcare Interoperability Resources (FHIR) are important for combining data and creating large sample sizes to study and understand results that examine diverse sub-populations. In turn, this reduces research bias due to the geographical and socioeconomic diversity of the larger sample size. Additionally, FHIR enables the use of health facilities' existing investments in certified electronic health record technology (CEHRT).

While HL7 FHIR Application Programming Interfaces (APIs) establish a solid foundation from which to promote consistency and interoperability, they are not a complete solution to establish rigorous quality and standardization required to stand up clinical trials in an emergency. Specifically:

- FHIR APIs provide a degree of standardization, but primarily as an interchange and API specification and not a clinical data model. This means that additional work would be required to transform the data into a suitable format for clinical informatics and observational research.
- Data quality issues will persist and vary due to the heterogenous nature of contributing sites.
- Variance in electronic health record (EHR) data capture across sites and different approaches to architecting FHIR representation can introduce subtle biases that should be understood and addressed.
- More stringent measures should be put in place to encourage the implementation of Privacy Enhancing Technologies (PETs) that will unlock the potential of large datasets and establish the framework to enable the swift deployment of clinical trials in during emergencies. Such a capability, empowered by data analytics and PETs, would benefit from expansion of information blocking regulations to cover sharing information that is not electronic health information (EHI). For example, it is unclear how an organization would train a machine learning model using federated learning on data accessible only through HL7 FHIR APIs/US Core APIs. API standards specified in the Code of Federal Regulations (CFR) § 170.215 were simply not designed with PETs in mind. Sharing broader data sets, even when not entirely necessary given the use case, can vastly increase the impact of analysis, and the power of that analysis to improve programmatic and organizational effectiveness.

While HL7 FHIR APIs provide a sound foundation for interoperability, significant downstream work is required to truly enable cross-site analysis possible in an emergency clinical trial. For example, researchers can leverage data quality metrics and dashboarding in a shared data repository to understand and characterize site-to-site variability and harmonization of lab units—including inference for missing units.

3. SMART on FHIR APIs

Palantir believes the portability of SMART applications would enable the Government to broadly reach and empower institutions with varying technical resources and competencies, as well as expand clinical research into underserved settings.

Specifically, the public health authority could create a SMART review and abstraction application and provide the application to healthcare facilities and institutions. This application could, among other things, enable the preparation of Case Report Forms (CRFs), including reviewing and correcting mappings of bulk FHIR data extracts and reviewing data completeness and quality issues flagged by the public health authority in the context of their EHR data. Such a capability would enable the public health authority to quickly adapt to the emergency as well as observations in facility and institution data.

4. Clinical Decision Support (CDS) Hooks

Building upon SMART on FHIR APIs, Palantir recommends that public health authorities and healthcare facilities / institutions review and update CFRs collaboratively, using a shared SMART application, and then push them into the clinical and emergency trial workflows using CDS Hooks.

However, this approach should be used sparingly, as it creates additional clinical documentation burden. Rather, non-standards based approaches for EHR-to-electronic data capture (EHR-to-EDC) should be considered as an augmentation to bulk FHIR. This approach is detailed further in Question 5 below.

5. Operationalizing Protocols of Varying Complexity

As established in our response to Questions 3 and 4, we firmly believe that bulk FHIR API exports can be used to effectively gather data for a protocol that is relatively close to the standard of care for a particular condition. FHIR questionnaires/questionnaire responses can be useful in capturing data for more complex protocols, however, this method should be a last resort. FHIR questionnaires should not be used when documentation requirements are duplicative of the incumbent EHR workflow. In those cases, FHIR questionnaires should be avoided in favor of non-standards based EHR-to-EDC data extractions.

Finally, only a small fraction of EHR data elements is required to be exposed by Certified EHR Technology via standard APIs according to § 170.215. Beingly solely dependent upon those standards (e.g., by using bulk FHIR exclusively) will severely limit data element availability that would be critical in an emergency clinical trial.

6. Consent, Deidentification, Return of Results

Consent and Authorization Management: Organizations working with sensitive health data must ensure fidelity to patient consent, including in situations where consent is granted and later retracted. To address patient privacy concerns, the ability to offer dynamic consent management in clinical trials will be a benchmark of success for research organizations. We believe that this capability will be particularly valuable for public and private sector organizations that are responsible for ensuring the comprehensive deletion—to the point of irretrievability—of emergency healthcare data, irrespective of how broadly it may have been shared in the course of legitimate use.

As COVID-19 approaches endemic levels and the emergency phase of pandemic subsides, researchers can no longer rely on emergency approvals for access to clinical data. As we move forward, trial operators must be able to speak to the consent status and use of individual patient data. An optimal data system will also allow them to show exactly how the data was used and, perhaps more importantly, prove that it was not misused. In complex systems, data can be replicated many times, combined with other data, and stored in different ways to support varied use cases, constituting the data lineage. When a data item needs to be deleted—e.g., in the case of a patient retracting consent—all instances of data use in the lineage must also be deleted in order to ensure that data is no longer accessible. This “hard deletion”—in which the data is no longer accessible anywhere in the system—is only possible with robust data lineage management tools. Such features include inherent auditing and version control through branching of data pipelines to trace both authoring of code and access to and use of data. Another layer of verification can be created through the use of metadata that describes how a given dataset relates to its parent and any other derivatives.

Effective consent management within a data platform should combine human input with software-backed checks, safeguards, download controls, and tiered access policies. For example, workflows for review and approval or publication requests can offer checkpoints to users for varying levels of consent. To ensure continued appropriate use, users can be required to justify downloads via prompt or confirm acknowledgement of an official policy or code of research conduct. Having a platform that identifies these layers of consent and flags them to users offers the opportunity to continually evaluate consent issues and even self-evaluate for violations against privacy laws or purely ethical considerations.

Deidentification: Palantir recommends a tiered, API-driven approach for deidentification of data in clinical trials. Using FHIR-compatible APIs will help streamline this process (see response Section 2 for further recommendations on these tools). In a large scale, emergency clinical trial setting, researchers need data to flow smoothly on a spectrum from more sensitive (e.g., Personal Health Information, or “PHI) to less sensitive (e.g., a de-identified dataset of vital sign readings) while staying synchronized.

Flexible and Responsible Reuse of Data: Clinical trials require a level of interoperability and modularity that allow organizations and users to interact with different providers and implementations of differential privacy. This infrastructure must also be flexible enough to be upgraded and replaced as new best practices are developed. The impact of clinical trials—especially those conducted in an emergency setting—is amplified when data is made available for reuse and collaboration. Organizations must balance the objective to create easy access to advance discovery with their legal and ethical responsibilities to patients, particularly in regard to privacy. Purpose-based access controls and the use of limited datasets can help strike this balance while driving research forward, whether study extends beyond the scope of a clinical trial or to engage more researchers toward a common goal.

Purpose-Based Access Controls: Purpose-based access controls play a critical role in fundamental trust in PETs as well as environments where PETs should be applied to protect data and promote participation. Such controls aim to:

- Introduce structure and clarity to data access decisions.
- Capture missing context and make it available to the people who need it.
- Build intuitive tooling for non-technical data governance teams to enforce rules.

Instead of applying for broad access to an individual data set, a potential user applies for access to a purpose. The purpose is set by data governance teams to contain data specifically scoped to help the user meet their goal.

Data governance teams must record a rationale for their decision at the same time they grant a user access to data. Likewise, data owners must record a rationale when they approve the use of a data set for a purpose. Recording these justifications prompts both sides to continually consider the necessity and proportionality of their decisions. The output of that assessment can be captured by software, making it available to data governance teams for review. At any point, an auditor can understand not just who has access to what data, but also why they were given access—with all the context that went into that decision.

Limited Data Sets and Safe Harbor: A key ability to foster collaboration around clinical trials is the use of limited data sets as defined by HIPAA. Under law, this information, while typically protected as PHI, may be disclosed to an outside party without a patient's authorization if certain conditions are met. First, the purpose of the disclosure may only be for research, public health, or health care operations, such as in the case of an emergency clinical trial. Second, the exchange of data must be governed by a data use agreement (DUA) that provides for the ability to audit compliance with the DUA as it outlines governance workflow around access to the data. Meanwhile, Safe Harbor standards include more restrictive privacy requirements that extend not only to individual patients, but also to their close family members and associates.

While these standards can be easily implemented by removing the appropriate HIPAA identifiers from a dataset, this process risks the loss of critical information related to temporality or granular geographical factors. In the case of complex, emergency clinical trials, an expert determination should be made about what set of privacy requirements and levels of access to data (as described in the above section "Purpose-Based Access Controls") are most appropriate to conduct the required work. This determination should weigh the need for protective access, audit, and review privileges against the urgency of the trial and need for results.

Machine Learning and Differential Privacy: Machine learning (ML) has opened increasing possibilities for the employment of differential privacy standards for research. A differential privacy approach ensures that raw data pertaining to individuals will not be viewed (and does not need to be modified) but instead is processed via algorithm. This algorithm effectively de-identifies the data by producing an output of insight that would not be significantly impacted by the input of one individual's data. When paired with a cloud system, this process can facilitate urgent research collaboration with massive, disparate data sets all while providing a guarantee that individual PHI is protected.

Return of Results: To encourage data sharing and remove the burden on organizations contributing data, the data infrastructure should be able to directly connect to source systems—as well as privacy preserving record linkage (PPRL) solutions—via bidirectional open APIs (including a FHIR-enabled APIs). Direct connection will not only remove reliance on intermediaries that aggregate and manipulate data and improve speed of access to and transparency into real-world data, but also offers more control and oversight to trial sites and other data contributors. In a clinical trial, these bidirectional connectors allow for an open flow of information back from the trial's data system to contributing organizations or participants, where appropriate, and with varying levels of deidentification and access. This allows participating institutions and organizations to benefit quickly and directly from the trial, incentivizing further collaboration.

7. User Interface and Experience

To optimize the experience of health care providers, administrators, and other users while maximizing the utility and update of the product, public health authorities should rely on EHR-to-EDC integrations to the greatest extent possible. This method should be prioritized beyond existing standards, particularly prior to requiring additional documentation burden for health care providers, administrators, and other users.

10. Emerging Technologies

The implementation of PETs offers opportunities for organizations to improve their management of critical, sensitive data (e.g., clinical trial data). Commercial software can integrate directly with PPRL PETs to securely link de-identified data across data sources. This can allow researchers to link clinical data from EHR records with additional information such as images, viral variants information mortality data, and Medicare claims. Such linking provides numerous benefits: analysis of x-ray images provides deeper insight into the impact of the disease on the lungs, variant data enables understanding of the clinical differences caused by different variants, mortality data promotes an accurate picture of patient outcomes, and claims data provides a complete picture of the medicines a patient is being prescribed. Access to such historically restricted data enables a richer understanding of disease while preserving the privacy of patients. Such PETs allow researchers to fill in previous gaps lost in the anonymization process, providing data once assumed unavailable that better reflects a holistic reality and is therefore more valuable. This model of PPRL linking has been successfully employed by Palantir in its work with the NIH's N3C, which facilitated the rapid integration, harmonization, and sharing of clinical data to provide a global, evolving view of the COVID-19 outbreak in the United States.

Integrated with other PETs as well as other technical products or governance procedures, PETs produce a contextually configurable, holistic data governance arsenal that can support robust security needs alongside complex access controls. However, like other emerging technologies, PETs rely on a strong foundation of accurate data to function optimally. Managing data governance and hygiene across disparate inputs is a prerequisite challenge for their successful implementation. While PETs can offer powerful capabilities as individual tools, they are most optimally deployed as part of an approach that addresses the full ecosystem and lifecycle of data management in complex real-world systems. In this more holistic setting, privacy risks are better addressed through a combination of several interrelated and reinforcing technological safeguards. Successful organizations will leverage a suite of intentionally redundant and interoperable PETs to enhance security at all layers of software and data architecture.

AI/ML also represent rapidly growing fields in emerging technology with the potential to transform clinical trials. While the application of these capabilities to clinical trials infrastructure has not been fully explored, its most promising and attainable application in its current state will be for operation and infrastructure. For example, algorithms may help to identify optimal size for a trial, target locations and populations, and improve overall efficiency of the process, preventing unnecessary spend and effort. As they continue to facilitate more applications of differential privacy, AI/ML may also transform standing clinical trial protocols for consent, relieving the burden of deidentification. Because the output of these algorithms embeds a privacy guarantee by

nature, further development in this space may eliminate the need for a deidentification process altogether as input data will never need to be accessed or viewed.

11. Pilot or Demonstration Project

A pilot or demonstration project should explore how the U.S. Government can proactively establish an infrastructure for emergency clinical trials and deployment that capability for a specific, example emergency scenario. The pilot(s) should mimic the real world environment to the greatest extent possible. The foundation of the pilot, and the long-term capability, should be managed through a logically shared/federated data repository (LS/FDR). A LS/FDR could operate as a shared cloud-based platform/infrastructure—providing central capabilities to researchers—while the actual data and access could be granularly configured to ensure data owners retain control and transparency over their data. Central, shared capabilities of an LS/FDR could include the SMART on FHIR application to validate data quality and standard adherence, and CDS Hooks to push standards' updates to healthcare facilities through the logically federated model.

OSTP's pilot(s) should explore the feasibility of public-private partnerships to enable this capability. Throughout the pilot project(s), OSTP should test the central repository's simplicity of design as well as the flexibility of the solution. A cloud-based solution, or option, will likely empower this initiative with the most flexibility and cost-effectiveness.

Additionally, the LS/FDR should demonstrate a variety of foundational capabilities, to include:

- *Interoperability*: Proven ability to leverage and integrate ONC's interoperability standards and technology, including USCDI and FHIR, as well as to harmonize data and common data models and terms (PCORnet, OMOP, HL7 FHIR, mCODE, etc.)
- *Data Integration*: Data integration must be a core capability of the piloted data repository. The capability should provide a highly configurable set of data integration tools that extend far beyond typical extract-transform-load (ETL) or extract-load-transform (ELT) solutions extending across data connections, transformations, and pipeline management. The data repository should be capable of integrating directly with PPRL providers to securely link de-identified data across data sources. Examples of PPRL providers include the Cancer Imaging Archive (TCIA) and the Regenstrief Institute, which enable linkage of clinical and imaging data.
- *Model Integration*: Due to the varied nature of anticipated emergency clinical trials, the data repository must be flexible and accommodate models created within the repository, imported from external environments, or configured as externally hosted APIs
- *Ontology*: The data repository should include an ontology capability. An ontology enables the link between digital assets and real-world counterparts and would enable the repository to aggregate and integrate data across sites using heterogeneous data formats or frameworks. Ultimately, it would facilitate the compiling of a larger clinical trial data set that is statistically significant and representative of a diverse population.
- *Analytics*: Users should have access to a variety of analytics tools for technical and non-technical tools alike. The central repository should be interoperable with common systems and analytical tools used by the research community.
- *Security*: Any data repository must be backed by rigorous platform security (granular controls, transparency, usability), enterprise security (mandatory encryption, strong authentication, security audit logging, etc.), and infrastructure security (security baseline configurations, robust security architecture, security monitoring, etc.).
- *Administration*: The data repository should include tools to facilitate the administration of robust clinical trial data. This includes configuring different views of Foundry for different users, authorization/authentication, and resource management that provides visibility into the utilization of the data repository resources.

Due to the sensitive nature of clinical trials, specific attention must be given towards the privacy rights of individuals who have agreed to participate in emergency clinical trials. As such, the central repository should provide a variety of inherent privacy enhancing capabilities including but not limited to:

- *Access Permissions*: Ensuring that users only have access to precise subsets of data necessary for their responsibilities.
- *Action Permissions*: Restricting permissions to conduct potentially sensitive actions, such as importing, exporting, transferring, or combining data to those users who absolutely need to do so.
- *Marking Data*: Persistently tagging sensitive datasets to clearly indicate their sensitivity, and to restrict actions such as joining them with datasets bearing other markings that may be risky in combination.
- *Obfuscation by Default*: Making data encrypted and unreadable by default. Users must enter an acceptable justification in order to decrypt necessary subsets of the data.
- *Auditing*: Empowering oversight bodies to check and verify compliance with data governance policies around de-identified data, and that no spurious, malicious, or risky actions are undertaken.
- *"Inferring" Sensitive Data*: Running background checks to infer sensitive data across the system, automatically flagging and locking down sensitive data uploaded accidentally or de-identified insufficiently.
- *Testing & Validation*: Providing the ability to do validations and "battle-test" anonymized data before it is shared more widely within the system or exported for external use.
- *Data Lineage*: Leveraging lineage tracking to understand how data is flowing within the system: which users have access to what level of identifiable data, and for what purposes at different stages.

OSTP should consider conducting a separate pilot to explore governance of the emergency clinical trial capability. Governance and a lack of standardization is likely to act as a significant bottleneck to any operational emergency clinical trial solution. By conducting a governance-focused pilot, OSTP could explore governance and policy that would simplify and expedite the deployment of clinical trials in real world emergency scenarios and better understand policy frameworks that will encourage participation and improve clinical trial outcomes.

12. Specific Commercial Capabilities

Foundry: Rapid Data Infrastructure Configuration and Deliverability for Emergency Clinical Trials and Health Research

Foundry is a commercial data integration and management platform that supports critical research across government health organizations and leading commercial life science companies. Foundry powers fundamental and translational research for the National Institutes of Health (NIH); and it's relied upon by providers, healthcare organizations, and healthcare systems such as the UK's National Health Service (NHS). In addition to providing research infrastructure, Foundry has also been deployed by the U.S. Government for high-risk emergency needs (e.g., disaster relief, disease outbreak) to provide leaders and operators with accurate, comprehensive data to prepare, mitigate, and respond to public health crises. Commercially, Foundry is used across the healthcare and life sciences value chain at leading companies like Sanofi and Syntropy for preclinical research, drug discovery, clinical trial site selection and harmonization, and cell line development.

To enable these complex missions and systems, interoperability and openness are central tenets of the Foundry at every layer. Foundry offers a variety of capabilities for interoperability such as integration of disparate data types, storage of data and logic in open formats, and access to transformation logic for replicability. Foundry has powerful capabilities for connection to systems and export of data through open APIs and bi-directional connectors; its Data Connection application includes over 200 out-of-the-box connectors for integrating with system types including SAS, SAP, Hive, Teradata, Sybase, DB2, Oracle, SQL server, FTP server, and HDFS. It also includes tools for building custom connectors, meaning integration options are not limited.

Simultaneously, the software's access control system and security features enable use at all classification levels and storage and protection of sensitive data such as PHI and Personal Identifiable Information (PII). Foundry allows researchers at various organizations to integrate, visualize, analyze, and leverage complex data in real time. With numerous applications and configuration options, Foundry's open architecture can be extended and modified in response to the ever-evolving needs of researchers and other users as they conduct emergency clinical trials or respond to rapidly developing public health crises.

Features and Capabilities

Foundry consists of hundreds of distinct services that cover a wide range of functionality. Together, these services combine to form a modular, end-to-end operations platform with minimal configuration. Interoperability and openness are key principles of Foundry as described in user documentation here. We offer a brief summary of platform capabilities below.

Flexible, Rapid Data Integration: Foundry offers several methods to connect to external data sources including out of the box connectors, open APIs, batch/stream connectors, and tools to configure connectors. Pipeline Builder is Foundry's primary application for data integration. Users can employ Pipeline Builder to build data integration pipelines that transform raw data sources into clean outputs ready for further analysis. Foundry can rapidly integrate new data of any source (e.g., registry data currently held in a Postgres database), format, or size. As a data-agnostic platform, there are no limitations to the type, size or number of data sources Foundry can integrate. Foundry can connect directly to any underlying source systems via a flexible set of adapters. This includes electronic health record (EHR) systems, Oracle, MS SQL Server, Azure, AWS S3, Cloudera, HDFS, REST APIs, SAP, SFTP servers, and more. An overview of the most common data source systems Foundry users integrate with can be found here. Because of its inherent flexibility, Foundry is adaptable to changing requirements, including the ability to support new data and metadata types, and new data and user scale. In a clinical trial setting, this provides an open infrastructure for lab reporting that both improves the quality and breadth of information collected and accessibility for organizations to accelerate discovery.

Powerful Analytical Tools: Out of the box, Foundry contains both point-and-click and code-based tools that enable table-based analysis, top-down visual analysis, geospatial analysis, temporal analysis, and more. For example, researchers of varying technical ability can collaborate together in real time to quickly iterate on exclusion and inclusion criteria to define a patient population for further study. Moreover, analytics in Foundry are designed to go beyond conventional "read-only" paradigms to write data back into the Ontology, producing valuable new insights within unified security, lineage, and governance models. Foundry's core Analytics applications include:

- *Contour:* A top-down analysis application for rapidly exploring tabular data at scale, deriving new datasets through visual transformations, and creating charts.
- *Quiver:* A multimodal charting application that allows for object-driven analysis, time series-driven analysis, point-and-click machine learning, and dashboard building.
- *Code Workbook:* An application that blends data engineering and data science motifs, allowing for Python-, R-, or SQL-driven transformations to be rapidly constructed, building and training machine learning models, and much more.
- *Notepad:* An integrated solution for embedding dynamic analytical, visual, and operational artifacts from across Foundry, alongside formatted text and media.
- *Fusion:* A spreadsheet-driven application that synthesizes tabular computation with the power of Foundry's Ontology and object-driven query system.

Security, Access Controls, and Sensitive Data Protection*:* Foundry provides tools to build a trusted research environment, with full transparency and controls over when, how, and why data is used in research. Foundry's extensive suite of security services (overviewed here) are designed to interoperate with customer and

organizational security systems and standards by leveraging existing authentication systems for identity, and existing authorization systems for permissions that can span role-based, classification-based, and purpose-based regimes. When handling patient/clinical data, the system offers sensitivity markings for known PHI datasets and rule-based alerts to proactively identify potential PHI in pipelines. Export of sensitive data can be gated, requiring justification and approval prior to download. Finally, access and controls can be modified at a granular level, ensuring that only the right users have access to the right information for a justified purpose.

***Data Discoverability**:* The Data Catalog is interactive view of curated data in Foundry. Depending on their permissions, users analyze every dataset Foundry. However, Data Catalog maximizes data discoverability, allowing researchers—even those with non-technical backgrounds—to see, filter, and analyze the data most useful and relevant to their study more quickly. The Data Catalog is organized into Collections, which contain curated data for a given topic. Foundry's robust indexing and search features, backed by Postgres and ElasticSearch, allow users to search broadly across their entire enterprise data landscape or filter for more targeted results. Data in Foundry is also stored in open formats, ensuring that information can be exported for use in external environments and applications. The storage layer uses open data formats (Parquet, Avro, etc.) that are 100% compatible with industry-standard tools (Spark, Flink, etc.). Foundry also ensures all data has its schema embedded within it, meaning that it is self-describing and does not require any proprietary content to make it usable or sortable, promoting easier discoverability for collaboration.

***Automatic Data Quality Checks**:* To ensure that data and metadata are continuously high quality and AI-ready, Foundry offers a set of configurable, automated data health tools and other QA/QC capabilities out of the box. All incoming data must pass rigorous data quality and privacy checks. If new data does not meet expectations, Foundry alerts administrators, who can review and resolve issues either at the source (i.e., with the data owner) or in pipelines. More information can be found about Foundry's Data Health service can be found here.

***Ontology**:* Once data is integrated, Foundry offers a common ontology that maps data to real-world concepts already familiar to users (e.g., "patient" or "lab facility"), links data between systems, and interoperates with standard health data models and taxonomies (e.g., FHIR, LOINC, SNOMED, USCDI). Foundry can quickly integrate disparate data models from across partners into a scalable, uniform ontology that can easily adjust over time. Unlike traditional systems, the Palantir approach to ontologies can encompass multiple data models side-by-side to support workflow and domain-specific needs. With the ontology in place, all elements can be accessed through REST APIs, facilitating synchronization with existing modeling tools and ontologies within data catalogs.

***Data Harmonization and Mapping**:* Foundry converges incoming data into a single model that incorporates desired data standards (e.g., CDEs) and any connected dictionaries, repositories, or catalogs. Data cleaning and pipelining is done using open coding languages (such as Python, R, and SQL) or intuitive, no-/low-code tools, enabling users to transform and map data according to a configurable OMOP or other standard models. This includes defining metadata schemas, data redaction/deidentification, and validations for consistent harmonization and modeling.

***Flexible Backing Stores**:* Data in Foundry can be written to multiple backing locations and can be stored on any Hadoop-compatible file store. Foundry can be easily configured to connect to existing clusters, read from those file stores, and write to them. When pulling from legacy systems that store data on non-Hadoop-compatible file stores, the built-in data connection tool can be configured to read from any source system, enabling easy data ingest or full migration.

Interoperability Use Cases Across in Healthcare

***National Covid Cohort Collaborative (N3C)**:* As the N3C Data Enclave, Foundry provides the ability to integrate patient-level data from a variety of different data models (e.g., FHIR, TriNetX, PCORnet and ACT/i2b2) from 77 different academic medical centers. Configuration of the mapping logic to a common data

model (OMOP) was completed in a few weeks by a diverse team working across multiple organizations and was made possible by Foundry's collaboration capabilities, including branching and versioning tools, and deidentification pipelines that include date shifting and ZIP code truncation. Mapping logic for data ingestion and harmonization and certain data pipelines have been templatized and made available publicly via the N3C GitHub repository, allowing logic to be applied instantly to data from new sites. This project represented a major shift in how data sharing is done in academics from a federated model to a hybrid model where rapid investigation can be done, setting the stage for future collaborations.

NIH Integrated Data Analysis Platform (NIDAP)*:* For over three years, Foundry has hosted and powered NIDAP, enabling effective collaboration and analysis of biomedical R&D data at scale, including data from disparate internal and external sources. NIDAP centralizes research at the National Cancer Institute, offering secure access controls, robust analytical applications for biomedical research, visualization tools, and collaborative tooling that can be used with no technical background.

Sanofi: Sanofi, one of the world's leading pharmaceutical companies, has used Foundry to provide core data infrastructure and an analytical platform as part of their platform for real world evidence (RWE) research. Foundry is used to integrate medical data sources for over 300 million patients into a common foundation totaling tens of billions of medical records and many types of structured and unstructured data (e.g., claims, electronic medical records, patient registries, wearables). In addition to facilitating sophisticated population analysis at scale, Foundry also allows the organization to run studies in collaboration with a range of internal and external Sanofi partners. Foundry provides the governance tools required to ensure that data is used appropriately and in accordance with relevant regulations and usage agreements. Its data management tools ensure that data is continuously updated as new real-world data becomes available. Lastly, its suite of powerful analytical tools allows Sanofi researchers to implement sophisticated epidemiological studies using statistics and machine learning.

Clinical Stage Biotechnology Company*:* At a leading clinical-stage biotechnology company, the translational research team automates complex clinical trial analyses using Foundry for faster therapeutic discoveries. Diverse datasets and AI/ML models in Foundry provide critical context on the origins of patient response differences. Such capabilities enable this R&D driven organization to make more informed decisions about which patients will most benefit from their first cell therapy programs and enable effective prioritization of future research efforts to bring novel cell therapy treatments to patients.

Syntropy*:* In a unique collaboration between EMD Digital Inc., an affiliate of Merck KGaA, Darmstadt, Germany, and Palantir, Syntropy utilizes Foundry to advance research by providing a trusted environment for scientists and institutions to manage the major challenges of data management, governance, stewardship, and data sharing. At major scientific research institutions such as MD Anderson and the University of California Irvine, Syntropy utilizes Foundry's technology to help researchers to contextually integrate and draw clinically meaningful insights from vast quantities of data, including clinical, biospecimen, imaging, and other sources. Syntropy also assures appropriate use and data protections through state-of-the-art provenance and access controls. By assembling and harmonizing the diverse data types and making them "similar" enough to analyze in a digital ecosystem while highlighting their unique differences, researchers are able to access data, establish collaborations, and team together to efficiently derive insights more effectively.

HHS Protect*:* Foundry serves as the data ecosystem for HHS Protect, interoperating with other commercial technologies to power the COVID-19 response. HHS Protect integrated over 200 data sources from all 50 states as well as data from over 6,000 hospitals, all U.S. Government public health labs, over 80% of U.S. private labs, and the Indian Health Service to support analysis, modelling, and information sharing in the fight against COVID-19.

Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

January 27, 2023

https://www.federalregister.gov/documents/2022/10/28/2022-23489/request-for-information-on-data-collection-for-emergency-clinical-trials-and-interoperability-pilot

The following comments have been submitted on behalf of the *Vulcan HL7 FHIR Accelerator Program*. Questions and clarifications can be obtained through contact at vulcan@hl7.org.

Vulcan is part of the HL7 FHIR Accelerator Program, focusing on the development of interoperability standards for clinical and translational research. Vulcan brings over 40 organizations together to provide an open, transparent and non-biased community for standards development and implementation.

The current program includes the development of Implementation Guides / standards for:

- ***Real Word Data (RWD):*** Extract data from EHRs in a standardized format to support clinical research and especially submission to Regulators
- ***Schedule of Activities (SOA):*** Represent the schedule of activities in FHIR from a spreadsheet. Enable the consistent description, timing and identification of each activity in a study
- ***Phenotypic Data:*** To increase the availability of high-quality standardized phenotypic information for genomic research and genomic medicine
- ***Electronic Product Information (ePI)*** : Define a common structure for product information (monographs) that supports cross-border exchange of data for patients
- ***Adverse Events (AE):*** Support standardizing the reporting and format of an adverse event. Improve the maturity of the relevant FHIR resources
- ***FHIR to OMOP:*** Support the development of FHIR to OMOP data transfer for better analysis of clinical data for research

RWD, ePI and SOA are currently being balloted by HL7 and we expect these to be declared as STU (Standard for Trial Use) in the coming months. Vulcan is also actively pursuing additional use cases / projects to support the research community in the coming months, as well as implementation of our STU standards through pilots, proof of concepts and real-world implementations. Vulcan leverages HL7 Connectathons to validate the standards developed by our members.

We are happy to provide comments through this RFI process and welcome the opportunity to work with OSTP and ONC on the pilot program. We believe that Vulcan can play an important role in community outreach, structure and governance to make your program effective.

Part 1 (of 2): Use Case Comments

OSTP is still in the process of collecting information on governance models and other aspects of the emergency clinical trials initiative. For purposes of responding to this RFI, however, we would like responders to consider the following multi-step use case.

For the purposes of this RFI, we are interested in the feasibility of all steps in the *(below)* hypothetical use case; we would also like input on how much of the use case could be operationalized in a pilot or demonstration project that might move forward in a timeframe of 6-12 months from the close of comments on this RFI.

Use Case Steps	Official Feedback

1. A U.S.-level governing entity would oversee development of a clinical trial protocol for broad distribution across clinical trial networks and sites.	• Vulcan creates HL7 FHIR Implementation Guides (IGs) to support clinical trials. There are 3 Vulcan Implementation Guides completing ballot for Standard for Trial Use (STU):
	  ○ Schedule of Activities
	  ○ Real World Data
	  ○ Electronic Product Information
	• Vulcan invites OSTP / ONC to co-develop future IGs and to adopt the existing IGs, such as Vulcan Schedule of Activities (SOA) for execution of protocol design, to leverage the interpretation of HL7 FHIR in clinical trial protocol.
	• Additionally, Vulcan invites OSTP / ONC to work with existing Vulcan members (e.g., HIT & EHR vendors) to define the strategy to determine how to implement the digitized clinical trial protocol into existing workflows and its access by supporting research networks.

2. Study sites would enroll participants in the trial (potentially using software mechanisms that can alert sites to potential subjects for a specific protocol in a manner that increases the diversity of trial populations). Sites would obtain appropriate e-consents and authorizations from participants.	• Utilizing a digitized protocol can provide a mechanism to use FHIR-based notifications and trigger conditions where potential participants are made aware—furthermore, the process where greater outreach in more diverse geographic locations and sites can be achievable. For example, the subject eligibility work within the HL7 Biomedical Research & Regulation (BR&R) Working Group is defining the FHIR requirements and demonstrates how patients can be engaged concerning clinical trial opportunities. • Helping diversify research sites to accommodate the challenges encountered during the COVID-19 pandemic. For example, using FHIR to determine healthcare site resources and qualifications (e.g., specialties, number of oxygen ventilation units, ICU beds, etc.) in correlation to clinical trial protocol requirements. • An ability to establish feasibility data in a coordinated and interoperable manner with HCPs, sponsors, among others. For example, using the USCDI as the basis to further identify data elements within FHIR to achieve inclusion criteria screening.
3. Clinical trial data is typically sent to the trial sponsor though an electronic	• There's an opportunity to shift from a typical to a more modernized methodology to collect and exchange clinical trial data. This reconceptualization involves using FHIR as the backbone of the collected clinical trial data, which is highly reusable, standardized, semantically decipherable, and adapted for human and machine

Use Case Steps	Official Feedback
case report form (eCRF), which is the record of data that is required under the protocol to be captured for each trial participant. A data element in an eCRF is the smallest unit of observation for a particular subject.	applications. For example, the FHIR Source Data Capture (SDC) uses Questionnaires as the basis to represent structured content similar to how traditional eCRFs work. In addition, the Vulcan Adverse Event (AE) is also evaluating how to transform an FDA MedWatch form into an FHIR representation to exchange data. • The work involved with this aspect can further enable the evolution of the minimal set of FHIR content using the USCDI as the foundation. In addition, however, the focus of this activity can help further define the connections between clinical care and clinical research, fostering further interoperability and eliminating redundancy and inefficiencies. • Reconfigure the way that data is collected from the point of care and reuse the existing data exchanged to Clinical Research utilizing the FHIR standard. Collect additional data elements needed for research in the most efficient, streamlined way to fit into the clinicians' current workflows.
4. The eCRFs would be transmitted electronically via common APIs to the sponsor.	• FHIR has an enabling infrastructure using web-based technologies to exchange data using APIs. Furthermore, with the mentioned SDC, SMART-on-FHIR applications, Clinical Quality Language (CQL), and FHIR Bulk Data Access it provides the mechanisms and utilities to fulfill interoperable data fluidity. The testing of said components can occur via FHIR Connectathons, where Vulcan can lead the design and development of those activities.
5. The study site's health IT system would present the eCRF content to clinicians in a manner that expedites data collection and (ideally) fits within clinician workflows.	• FHIR can leverage a mesh network of conforming servers to help eliminate data redundancy, adhere to proper data provenance principles, and more efficient clinician and administrative workflows/collection (e.g., the reduction or elimination of manual transcription, repopulation of repeated data elements, etc.) • Vulcan can identify additional common, standardized data elements that cover traditional eCRFs (e.g., via FHIR SDC or equivalent). In addition, Vulcan's membership includes multiple research community perspectives that include regulators, clinicians, vendors, and professional organizations. • Vulcan will also ensure caution about *mandating* certain data elements as that data may not be available / collected during the clinical workflow. Historically, the experience has been that requiring the data creates unintended consequences that impact completing the data elements and/or participating in totality. Such details will be reviewed and discussed in a collaborative community setting. • Other applications, such as SMART on FHIR and CDS Hooks, help clinicians consistently capture data and obtain knowledgeable insights across multiple systems.
6. As the clinician obtains data elements to complete the eCRF, that data would be captured in the patient's electronic health record.	With regards to this use case, there are several things to keep in mind. • Not all data collected as part of a clinical trial is collected within an eCRF (e.g., device data, mHealth data, central lab data). In this case, if such data is required to be made available in an EHR, this can be a direct source to EHR connection using APIs or other standardized and pre-existing methods. • Not all data collected as part of a clinical trial will be necessary to collect in an EHR (may not be relevant in a clinical practice setting).

Use Case Steps	Official Feedback
	• Permission must be granted by a study participant before data is pushed from study source system(s) to the EHR. Study participants are agreeing to be a part of the study, but that does not imply that they want their data to be linked to their health data in their medical record. • Study participants likely seek care from various healthcare organizations and can be seen at multiple, unrelated facilities, which means that they could be associated with multiple EHRs. This will need to be accounted for and decisions made on how to handle sharing between study system(s) and the various EHRs. • For multi-directional communication to be successful, data linkage is critical to ensure that the correct data is entered into the correct chart. Data linkage is not trivial and will require thoughtful consideration on the most appropriate methods given the data source and EHRs involved. While FHIR can facilitate the transmission of data from point A to point B, the data linkage process must be determined first. Additional comments: • Considering privacy and security requirements, the data produced from clinical trials (e.g., observations, diagnostic reporting, etc.) can benefit the patient's health records. Using FHIR can provide the benefit of identifying which data endpoints can be captured as part of the patient's health profile and exchange them in a secure, compliant manner. • Careful review of the data dictionaries for studies – of the intended definitions for each variable – is critical so that there is a clear understanding of what is being collected and how it best fits into the EHR. Often times, terminology used in a trial does not always translate to the EHR-specific definitions. • There is a need for systems to identify observations / diagnostics that are part of a trial (entered via an eCRF) as opposed to part of general clinical care. A standardized extension to the FHIR Encounter resource should be considered. Adding information back to the EHR needs to be done with caution (e.g. identify source of data, terminology mis-match, etc.). Note: may also require extensions for research data outside the existing FHIR resources (e.g. lab result research data). • Completion of the data elements required for a study includes patient reported data that can be documented by the clinician and/or the patient (with appropriate identification of the source of the information). • Other sources of clinically relevant data may exist outside the EHR. Consideration for bi-directional flow of data entry into and out of the EHR.

Use Case Steps	Official Feedback
7. The clinical trial data would also be sent to a central data repository or small set of data repositories for researchers to analyze. *[snip]*	• The clinical trial data shared via FHIR to consuming research repositories is achievable using FHIR Bulk Data Access and API methodology. Moreover, the solution can further apply patient tokenization technologies (with appropriate consent) to share data with applied privacy concerns taken into consideration.

Part 2 (of 2): Topic Comments

Our goal for this RFI is to support optimized data collection for clinical trials carried out across a range of institutions and sites, both in emergency settings and in the pre-emergency phase, under the use case described above. We also seek input specifically on the value of designing a pilot or demonstration project to operationalize data capture in the near term, for example within 6-12 months of the close of comments on this RFI. With those goals in mind, we request input on the following topics:

Topic	Official Response
1. United States Core Data for Interoperability (USCDI). We seek input on how U.S. Government and external stakeholders might leverage USCDI and future extensions of USCDI standards (such as USCDI+, an extension that supports federal partner program-specific requirements) to support emergency clinical trial research. *[snip]*	• Existing USCDI data classes and related FHIR resources provide a starting point for capturing trial-related data including patient demographics, vitals, test results, medications, procedures, and conditions. • Additions to USCDI that would benefit clinical trial data collection include the Consent, ResearchStudy, ResearchSubject, and AdverseEvent FHIR resources. • Updating USCDI to include the data elements required for research should not wait for an emergency but should be included now to benefit all research. • Several Vulcan projects have done a gap analysis between USCDI and the data required for their use case – these analyses should be considered when deciding on additions to USCDI. For example: Adding a field to the Encounter resource to distinguish those encounters in an EHR that are for patient care vs. those encounter records that form part of the protocol. • The existing Vulcan projects can help support the additional use cases to facilitate this process.

Topic	Official Response
2. HL7 FHIR APIs. We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-emergency phase, and in what areas further advances might be needed. Specific topics in this connection include: *[snip]*	• FHIR Bulk Data Access will be essential in the transmittal of large datasets. However, FHIR Bulk Data Access isn't necessary for all use cases proposed in this RFI. As offered, it is a transactional use case (seeing patients in a clinical workflow and capturing additional supplemental data); perhaps long-term follow-up might be a use case for bulk – e.g., once patients aren't being seen for the purposes of this trial any longer, but that's not outlined as part of this use case. Suggest deferring discussions of bulk FHIR for future consideration when a use case is better defined. • The Questionnaire / QuestionnaireResponse resources: FHIR resources already supported by certified EHRs and fit for purpose to the data classes of interest for the trial should be used. The FHIR Questionnaire resources should be reserved for data that isn't otherwise represented by the fit-for-data-class FHIR resources.

3. SMART on FHIR APIs: We seek input on how U.S. Government and external stakeholders might leverage SMART on FHIR APIs, and in what areas further extensions might be needed. It would be helpful to receive comments on: *[snip]*	• SMART on FHIR can help support emergency clinical trials and expand clinical research into underserved settings in several key ways: 1. Reduce the burden of research participation for institutions that have limited IT resources 2. Enable launch points from directly within the patient chart, providing a highly integrated way to collect research data with minimal disruption to existing clinical workflows. 3. Pull existing data from the EHR via the FHIR API to auto populate areas within the FHIR app and highlight areas for manual data entry 4. Write data to multiple points at the same time (e.g., to the EHR and simultaneously to a registry) 5. Standardize the requested data elements across all organizations 6. Provide links to educational content housed by government websites • Supporting wide adoption of automated research data collection technologies requires minimizing privacy, security and implementation risk to sites. • Stakeholders at each site must participate in, or approve, any integration with the EHR, including information technology (IT) technical personnel, IT project managers, information security, and legal. As most of these individuals are not directly involved in the trial itself, reducing burden and risk on hospital IT and security staff is critical to timely and successful implementations. • SMART on FHIR is designed to reduce these burdens and risks, and to be deployable across heterogeneous sites and trials in a highly uniform manner. It uses a recognized security model that requires no software or hardware installs and can be used to implement a user-driven and model for data transfers. This open source, standards-based integration framework is based on open standards including HL7's FHIR, OAuth2, and OpenID Connect and is supported by major EHR vendors with well-defined processes for validation, approval, and implementation. • A SMART on FHIR app could identity the data elements needed for trial participants, pull the available data points from the EHR and present them to the provider and highlight where more data is requested. For data points that are part of USCDI, mapping should be less complex and uniform across sites in the trial. • When the record is completed, data can be sent to the research database and, when needed, written directly back to the EHR. Additional features such as "how to" videos, protocol training resources, and patient educational materials could be included in the apps. • An organization like Vulcan, with membership that includes EHR vendors, the research community, and SMART on FHIR app developers, can help to define minimum standards for the design, functionality, and implementation and maintenance of SMART on FHIR apps. • Best practices for governance, IT enablement, and validation of data mappings per site & study, as well as support for interoperable digitized protocol definitions (as discussed in use case #1), are important to successful adoption and reuse across trials, sites, and sponsors.

Topic	Official Response
4. Clinical Decision Support (CDS) Hooks: We seek comments on how the HL7 CDS Hooks specification might be used to support clinical research, for example by creating prompts within the practitioner workflow during interaction with patients; and any advances that might be needed to support the use case described above.	• CDS Hooks could be explored to support clinical research in the area of recruitment, however this must be balanced to not create alert fatigue to clinicians, especially during an emergency. • CDS hooks might be useful for systems that can't do recruitment natively in their EHR but may not be required (or desired) for systems that support this already. • More discussions around the appropriate use cases for clinical research would be needed in an environment that brings the research community together to discuss HL7 standards, such as Vulcan.
5. Operationalizing protocols of varying complexity. As noted above, emergency clinical trial designs could range from relatively simple protocols to more complex studies involving the evaluation of investigational agents. We would appreciate comments on the following topics: *[snip]*	• Optimizing protocol design and development is needed. For example, TransCelerate's Common Protocol Template (CPT) helped define an initial foundation, with digitizing the protocol as the next step. • There are efforts underway to create *standard* international digitization of the clinical trial protocol. Vulcan offers the opportunity to participate in / support those external efforts to develop the standard. • Vulcan agrees that the primary long-term strategy is to leverage a FHIR-enabled clinical trial protocol for interventional studies that can be accessed and exchanged during emergencies and preferably for all research. The fundamentals of the process are flexible using the FHIR representation of the protocol (e.g., from PlanDefinition to an Observation). • For consideration, when linked with an electronic data capture system for research, a SMART on FHIR app can allow mid-study data definition and data mapping updates with audited publishing and change control - ensuring data integrity and traceability for research without burdening site research or IT staff with the need to make technical changes to the EHR or the integration layer. • Studies may benefit from the ability to launch complex, study-specific data collection instruments from the study electronic data capture (EDC) / direct data capture (DDC) system via SMART on FHIR and write the results back to the participant's health record potentially via FHIR Questionnaire / QuestionnaireResponse resources. • Care must be ensured to balance the capabilities of FHIR and systems with the workflow of clinicians and researchers to encourage adoption.

Topic	Official Response
6. Consent, deidentification, return of results. The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution. *[snip]*	• The approach to patient consent must be thoughtful. The ability to use existing eConsent technology is closely tied to the digitized protocol and recruitment / enrollment workflow. • Supporting the FHIR Consents resource so that regardless of whether a stand-alone eConsent strategy was used or a strategy native to the EHR, the metadata confirming the status / receipt of the signed consent for the specific study can be received from the sponsoring organization. • Development of the consent resource is an area that would be well-suited for Vulcan to assist.
7. User interface and experience. With all of the above technologies, we seek input on: *[snip]*	• Organizations, such as Vulcan, bring together the voice of the research community, health IT experts, regulators, patients, academia and health systems to provide guidance and / or recommendations on workflow design. • The success of any integration is dependent on fitting into the workflow and should not be defined by any single vendor, but rather a consensus amongst all affected stakeholder (e.g. researchers, providers, caregivers, and patients). • Vulcan brings together a large number of diverse organizations that each have a part to play in developing and implementing clinical research standards. Vulcan offer an open, consensus-driven and non-commercial based environment, resulting in broad-based standards that not only meet a broad audience, but also are devoid of a singular perspective.
8. *Capturing data elements required for clinical trial protocols. [snip]*	• Please see response to Part 1, Questions 1 & 3 and Part 2 Question 5.
9. TEFCA and QHINs. As noted above, TEFCA is in the implementation phase at this time. In the future, the TEFCA QHINs are expected to support implementation of the FHIR APIs (see the ONC Recognized Coordinating Entity's January 2022 FHIR Roadmap for TEFCA Exchange [6]). *[snip]*	• TEFCA and QHIN offers multiple opportunities for research, such as a centralized repository for feasibility, and could potentially benefit from the digitized protocol and FHIR-enabled workflows. We propose addressing TEFCA as a separate RFI.

Topic	Official Response
10. Emerging technologies. We welcome comments on any future technological developments we should anticipate. *[snip]*	• The use of patient, 3rd party, and 1st party tokenization solutions could facilitate many data privacy and security concerns. • The application of Artificial Intelligence / Machine Learning can benefit from the semiotics using the FHIR strategy (e.g., the semantic and syntactic components).
11. Pilot or demonstration project. We seek comment on how the U.S. Government can best work with external stakeholders and developers to develop a pilot or demonstration project that will operationalize clinical trial data capture and serve as a basis and model for data collection in the event of an emergency. *[snip]*	• Vulcan invites OSTP / ONC to utilize Vulcan's infrastructure, existing Vulcan projects (e.g., Real World Data, Schedule of Activities, Adverse Events, Phenotypic Data, electronic Product Information and FHIR to OMOP) and Vulcan members to design, develop, implement, and test minimal viable product demonstration projects and pilots. • Demonstration projects must be tightly focused to be achievable, using existing USCDI / FHIR capabilities with a defined timeframe, success measures and outcomes. • Consideration of alternative approaches, such as using patient facing apps and listening to the lessons learned and best practices from Vulcan's members' experiences (pilots, current production capabilities) would provide great insight. Moreover, the broader HL7 community can become a collaborative opportunity to mitigate the technology, standard, and specialized resources required. • Vulcan is currently exploring pilots / proof of concepts / implementation projects and is open to working with OSTP / ONC on this pilot initiative. In fact, we believe it's the best strategy.
12. Specific commercial capabilities. Commenters who are developing a technology or product that might be relevant to any of the topics set forth above are welcome to include a description of that product. *[snip]*	• Vulcan is dedicated to developing free and open source HL7 FHIR standards. Individual members of Vulcan are commercial, academic and government entities that have developed tools and standards to solve the research community's challenges. • Member companies share tools and products at the Vulcan Implementation Showcase. Vulcan invites OSTP / ONC to attend upcoming Vulcan Implementation Showcase, present, or review past showcase events for relevance.

DATAVANT

44 Montgomery St, 3rd Floor ♦ San Francisco, CA 94104

Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot
Office of Science and Technology Policy (OSTP)
submitted via email datacollectionforclinicaltrials@ostp.eop.gov

Organization: Datavant, Inc.
Respondent Type: Industry
Contact: Doug Fridsma, CMIO, doug@datavant.com

Organizational Details: Datavant is the leader in privacy preserving data exchange, working with over 500 institutions to connect health data. Our mission is to connect the world's health data to improve patient outcomes and bring new treatments to patients faster. To accomplish this, we are connecting a network of companies, non-profits, and government entities that utilize our common infrastructure for the safe exchange of patient-level health information.

At Datavant, we believe that data fragmentation is the largest challenge facing the health data industry, and protecting patient privacy is paramount when using health data to improve health and health care. We are focused on building an open data ecosystem that allows stakeholders in the healthcare system to freely exchange data while protecting patient privacy.

Datavant provides three key solutions within the health sector privacy-enhancing technologies (PET) space:

- Privacy-preserving record linkage (PPRL), which enables disparate records to be linked in a de-identified manner,
- Data de-identification and redaction tools and services, which enable data to be redacted and modified to meet the definition of de-identification within HIPAA,
- HIPAA Expert Determination and data risk disclosure tools and services, applying statistical and cryptographic expertise to ensure datasets formulated meet the definition of the HIPAA Privacy Rule for the Expert Determination Standard §164.514(b)(1).

Datavant's privacy-preserving record linkage and de-identification technology is a foundational, neutral privacy enhancing technology. It has been used to power innovative solutions that enable scientific advancement while preserving individual privacy. The use cases that this technology powers includes, but is not limited to:

- The formation of registries and data collaboratives such as the NIH National COVID Cohort Collaborative, N3C.

- Linkages between trial data and real world data sources (e.g. claims, EHR data) to form more complete longitudinal views of clinical trial cohorts for long term safety and effectiveness studies.
- Discovery of shared patient cohorts across disparate datasets to form more complete longitudinal medical records for patient cohorts of interest.
- Real World Data repositories to power large scale evidence generation studies.
- Linkages between data sources that fall under differing privacy frameworks such as health data and social determinants of health data.

We draw on our experience across all these various use cases in response to this request for information regarding advancing privacy-enhancing technologies.

USCDI and FHIR standards should serve as the starting point for clinical trials data collection

Throughout this RFI, there is the suggestion that the USCDI and FHIR standards should serve as the starting point for data collection. We agree that this is the right strategy and provides a mechanism to align clinical trials data with data drawn from EHRs. This will accelerate the use of real world data (RWD) in the use of clinical trials, and make the rapid deployment of clinical trials possible by leveraging EHR technology and existing FHIR and USCDI standards.

Where standards do not exist, these standards should be developed within a framework that allows them to be eventually incorporated into the USCDI and FHIR standards. This would include extensions to existing standards to support clinical trial specific use cases, and the experience of using these standards within emergency clinical trials will inform future standards.

Consent, de-identification and patient access

While most IRBs have concurred that PPRL generation and linking of trial data in a de-identified manner falls under Non-Human Subject Research, it is our posture that if it is feasible to collect explicit patient consent for linkage of data to real-world data, study investigators should do so. Language regarding the future linkage of real-world data to a patient's trial data can be added as part of the broader Informed Consent Form (ICF) for a trial, or be standalone. Standalone follow-on/linkage study consents can provide investigators and patients with additional flexibility to extend the use of data beyond the original study period. It is our experience from working with Life Sciences companies that these follow-on consents, when presented at study enrollment, gain high participation (i.e. in excess of 80%).

If direct consent is to be obtained, individuals at each site who will be responsible for obtaining informed consent should be trained on PPRL technology and be able to answer high-level questions on the intent of linkage of healthcare data. Additionally, it is important to streamline the site experience as much as possible, minimizing technology handoffs, training requirements and site PII handling/information security evaluations. Many eConsent platforms (e.g. Medidata, Medable, Science 37) have been developed that can consolidate the site experience of consent capture, PII management, and PPRL generation.

It is Datavant's experience in de-identifying linked results that combine prospectively collected trial data (eCRF, ePRO, IXR, wearables) and real-world data that an adjudicated statistical expert-led approach (under HIPAA's expert determination de-identification method) can preserve privacy in an enriched linked data set that allows investigators to focus on inclusion of critical data to support trial events and endpoints.

TEFCA is currently too immature to support emergency clinical trials. There are other networks that would provide faster, more reliable data collection.

We do not believe that TEFCA is ready to support emergency clinical trials. The technical specifications are based on an old standard that will be replaced by FHIR and more modern approaches to information exchange and uses a query model to retrieve information. Current standards for individual access and provider-to-provider exchange are based on document-centric standards that are outdated and not suitable for use in emergency clinical trials. While this technical approach has been used to support treatment and is intended in the future to support payment and operations (collectively, TPO), research use cases are not formally part of the TEFCA specifications, and there is significant additional work that will be needed to determine if the governance structures and technical specifications can efficiently and reliably support emergency clinical trials research. There are existing networks that can be leveraged now to support emergency clinical trials (PCORnet, TrinetX, etc) and these have more experience with the specific challenges of clinical trials research. Managing consent, identified record retrieval, leveraging de-identified RWD to augment existing data, and the challenge of post-marketing research in the face of rapid clinical trials have not been tested with TEFCA and likely will take too much time to be responsive to the needs of emergency clinical trials.

Emerging technologies: PPRL is no longer an emerging technology, but has widespread adoption across providers, researchers, and life sciences communities.

Privacy preserving record linkage (PPRL) solutions are no longer an emerging technology, but have widespread adoption across providers, researchers and life sciences. Datavant has now partnered with over 500 organizations (including most

leading EHR, claims, lab, pharmacy, and consumer data platforms) to comprise the industry's leading open, linkable Real-World Data ecosystem.

In clinical trials specifically, Datavant is now generating PPRLs for 80 clinical trials across 22 sponsors, with more than 110,000 participating patients and is now beginning enterprise relationships with life science companies who are generating PPRLs for most active trials. The net benefit of this growth is that sites, IRBs, investigators, and regulators are becoming increasingly knowledgeable and comfortable implementing these solutions. Using these record locators, sponsors intend to supplement the prospective evidence they are generating with data from the RWD ecosystem, to support assessments and endpoints based on their participants' complete medical history.

For example, we have direct experience in using PPRL in emergency clinical trials. A vaccine developer consented participants in a vaccine clinical trial to have their data tokenized and de-identified as part of their ongoing study. This proved to be fortuitous when an unexpected concern about cardiac arrhythmias associated with vaccination was raised. Using the de-identified tokens, they were able to rapidly identify past medical records for these patients, and link them to current study data and follow through. Rather than stop the trial or repeat it with specific questions related to cardiac arrhythmias, the investigators were able to identify previously unknown pre-existing conditions that explained the concern, and they were able to continue the trials without interruption.

Specific commercial capabilities

Trial Tokenization is a straightforward process that uses personally identifiable information (PII; e.g., first name, last name) on patients enrolled in a clinical trial to create a universal, de-identified token (also called a Privacy-Preserving Record Locator) that can be referenced to link records across datasets. . By tokenizing clinical trial data, life science companies gain the ability to link real-world data to their clinical trial data at any time, without unblinding the study or compromising the privacy of trial participants. Whether a company wants to link to electronic health records to run a hybrid trial, pull in diagnostic lab data for retrospective sub-cohort analysis, or look at claims data for long-term surveillance monitoring, it all starts with tokenizing trial data.

RWD data sources use Datavant software to de-identify their data and add unique linking tokens, but each source (and each clinical trial) is provisioned with a different encryption key. That means that the same patient will have different token values in each data set (or trial). This difference means that a breach at one site that somehow exposes the identity of a token will never propagate to any other site, because that token value will not exist at any other site. Only when two sources wish to allow their

data to be linked will the Datavant software convert the tokens into a common encryption key to allow matching.

Datavant's process for creating tokens emphasizes and maintains the principles of good clinical practice, and enables Datavant's partners to comply with corresponding regulations. The solution is designed to give patients' rights, safety, and wellbeing foremost importance. Token creation, by default, protects patient privacy, minimizes risk, and provides compliance with oversight and regulations.

Datavant tokens are created in clinical trials only when an institutional review board (IRB) or independent ethics committee (IEC) have given approval or favorable opinion or an exception.

Additionally, creating Datavant tokens upholds requirements for accurate reporting, maintenance, and verification of electronic records, including protections for patient confidentiality, in compliance with necessary regulations like HIPAA and 21 CFR Part 11. The site token-generation experience can be managed directly by Datavant with a 21 CFR Part 11 and SOC2-compliant web portal or in concert with eConsent/eClinical partners who already possess PII and manage patient consent.

We are grateful to be able to provide input into this RFI on data collection needs for emergency clinical trials and are happy to provide additional detail and input as needed. We would welcome the opportunity to participate in pilots to explore how best to leverage data in emergency clinical trials and look forward to ongoing efforts within OSTP to advance clinical research.

Sincerely,

Douglas B Fridsma, MD PHD
Chief Medical Informatics Officer
Datavant

Response to Emergency Clinical Trials Data Collection RFI

Jan 27, 2023

Submitted to *datacollectionforclinicaltrials@ostp.eop.gov* from:

NHLBI Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies (CONNECTS) Executive Committee, Steering Committee, and study PIs:

Committee Leadership:
Robert Harrington, MD, Stanford University; *Executive Committee Co-Chair*
Clyde Yancy, MD, MSc, Northwestern University Feinberg School of Medicine; *Steering Committee Chair*
Serpil Erzurum, MD, Cleveland Clinic; *Steering Committee Vice Chair*
Diane Nugent, MD, CHOC Children's Hospital; *Steering Committee Vice Chair*

Steering Committee Members and Study PIs (alphabetical order):
Gordon Bernard, MD, Vanderbilt University Medical Center
Samuel Morris Brown, MD, MS, Intermountain Healthcare/University of Utah
Clif Callaway, MD, PhD, University of Pittsburgh Medical Center
Sean Collins, MD, MSci, Vanderbilt University Medical Center
Mary Cushman, MD, University of Vermont
Mark Geraci, MD, University of Pittsburgh
Adit Ginde, MD, MPH, University of Colorado
Michelle Gong, MD, MS, Montefiore Medical Center
Judith Hochman, MD, NYU Langone Health
Nigel Key, MD, Univ of NC School of Medicine
Jerry Krishnan, MD, PhD, University of Illinois Chicago
Lisa LaVange, PhD, Univ of NC Gillings School of Global Public Health
Macky Neal, MD, University of Pittsburgh
Tracy Nolen, DrPH, RTI International
Thomas Ortel, MD, PhD, Duke School of Medicine
Paul Ridker, MD, Brigham and Women's Hospital
Wes Self, MD, MPH, Vanderbilt University
Matt Shotwell, PhD, Vanderbilt University
Sonia Thomas, DrPH, RTI International

Background:

The NHLBI Collaborating Network of Networks for Evaluating COVID-19 and Therapeutic Strategies **(CONNECTS**) program is a component of **Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV)**. The **overarching purpose of CONNECTS** is to test host-directed interventions for COVID-19 via rapid, efficient, collaborative adaptive platform trials aimed at helping to slow or halt disease progression and speed recovery. Since 2020 CONNECTS has established a clinical trial platform spanning outpatient, in-patient (including ICU), and convalescent care. CONNECTS tested/ is testing 18 different intervention strategies in adaptive clinical trials. To date, CONNECTS has enrolled over 6,400 participants at more than 300 clinical sites both individually and as members of 20+ networks, mostly from the US, but also includes sites in Spain, Mexico, Italy, Brazil, and South Africa (**ACTIV 4a, 4b, 4c, 4 HT and C3PO**, https://nhlbi-connects.org). Unusual for clinical trials, and in response to the disparities in COVID-19 infection and mortality, approximately fifty percent of participants are from a race or ethnicity under-represented in biomedical research. Most of the patients enrolled were hospitalized for COVID-19. The **strategic approach for CONNECTS** is to fully integrate existing NHLBI networks under one organizational umbrella to ensure efficiencies; standardization; collaboration; sharing of control groups (as appropriate), resources, and data, and nimbly shift studies as needed, based on new knowledge, and changing pandemic clinical landscape following an innovative model of seamless collaboration.

Non-government members of the **CONNECTS** Steering and Executive Committees and clinical trials PIs are responding to specific questions within this Request for Information on Data Collection for Emergency Clinical Trials based on the combined lessons learned from these trials.

5 *Operationalizing protocols of varying complexity.*

a. Any technical limitations that we should be aware of regarding use of the above tools to operationalize clinical trial protocols.

- Models to use EMR extraction to decrease site data entry would help alleviate work-force crush caused by the need for study coordinators to enter clinical trial data into an EDC system.
- This clinical trial data collection paradigm shift is in infancy and evolving yet needs to become the mainstay for all clinical trials to be effectively used in urgent settings requiring *fast* start-up from *many* sites.
- EMR solutions must be simple enough to implement technically such that community and rural care settings with limited IT staff resources and little experience in clinical research could implement them, as it is imperative to be able to add participants from this setting to studies of emerging pandemic illnesses.

- A demonstration project on CONNNECTS used REDCap software to extract patient specific trial data directly from the electronic medical record and place it into the study database. This was found to capture most of the data that was being collected by study coordinator direct entry into a clinical trial electronic data capture system, and potentially save time. This method depends heavily on setting up the mapping of the EMR to REDCAP at each site. Reference: Evaluating Automated Electronic Case Report Form Data Entry from Electronic Health Records | Journal of Clinical and Translational Science | Cambridge Core In order for this method to be effective, all sites, regardless of levels of technical support staff, would need to build standardized and consistent EMR data extraction for each clinical trial.
- Many trials require web tools beyond data entry provided in EDC systems – EMR data extraction would likely not replace the need for these tools (randomization, drug supply tracking, biosample tracking, SAE reporting, site monitoring, data queries/cleaning, etc.).
- Regulatory requirements for quality of the data for trials submitted to FDA for IND need to be developed.

-END-

January 27, 2023
Submitted via electronic mail

To: The Office of Science and Technology Policy
From: Verily Life Sciences, LLC
Re: The Office of Science and Technology Policy's Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot (FRN 2022-23489)

Verily Life Sciences thanks the Office of Science and Technology Policy (OSTP) and the Office of the National Coordinator (ONC) for the opportunity to provide a response to OSTP's Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot.

Verily is an Alphabet company whose purpose is to bring the promise of precision health to everyone, every day. Our work is focused on shifting the paradigm from "one size fits all" medicine to one focused on a more comprehensive view of the individual that leads to a more personalized path forward. We provide solutions across healthcare, from clinical research to care delivery, generating and applying evidence from a wide variety of inputs to change the way people manage their health and the way care is delivered.

In the clinical trials space, Verily has developed software that improves the research experience for participants, sponsors, and study sites alike. Using these tools, we are building disease-specific longitudinal registries that will provide deep insight on participant health and answer questions about which treatments work, and for whom they work best. Today, we work on registries and other clinical studies with organizations such as the American Heart Association, the Crohn's and Colitis Foundation and other leading life science and academic partners.

We provide comments on RFI questions #10a, 10b and 12 below.

10. **Emerging technologies. We welcome comments on any future technological developments we should anticipate.**

a. How future technologies might affect the use case and underlying assumptions laid out in this RFI.

One assumption of the use case appears to be that study data would be derived entirely from the eCRF, requiring a specific action by the clinician to input data into the eCRF. While this is a traditional method of study data collection, reliance on an eCRF alone for

study data may have significant limitations from an operational and scientific perspective. Notably, eCRF entries may disrupt the clinician's workflow; this can be mitigated to some extent by integrating the eCRF into the existing workflow (as described in the use case), but some disruption will likely still occur.

An emergency clinical trial infrastructure should seek to utilize data that is already captured in the electronic health record (EHR) and through real world data (RWD) sources. These data may be relevant to study outcomes and can meet U.S. Food and Drug Administration (FDA) guidance on the use of RWD for regulatory submissions, as appropriate. In addition, relying solely on data supplied by the clinician appears to preclude the use of patient-generated data that can be used to understand how a patient feels or functions during daily life and between visits to the study site through the use of scientifically validated tools (e.g., digital tools or more traditional survey tools).

An emergency clinical trial infrastructure should evaluate pragmatic approaches to collect relevant patient-generated data (e.g., regarding adverse reactions) that can be used to address scientific questions that are of importance to regulatory and public health decision makers, and that can increase trust from the public in the patient-centricity of the emergency clinical trial.

b. How future technologies might change the nature of the software architecture, data architecture, or potential data collection solutions for clinical trials.

Software architecture that supports emergency clinical trial infrastructure should incorporate the ability to collect, organize, and analyze significantly diverse sources of data including, but not limited to, the eCRF–such as EHR data, patient-generated data from digital tools and other sources, genomic or other biological data, etc. This data architecture should be able to address data quality questions (e.g., the confidence in linkage between data sets) that are important to regulatory reviewers and other public health decision makers. It should also incorporate robust privacy and security controls.

12. Specific commercial capabilities. Commenters who are developing a technology or product that might be relevant to any of the topics set forth above are welcome to include a description of that product. Comments about a specific technology or product should be limited to three pages or less.

A significant number of technology-enabled tools exist to perform clinical trial functions; SignalPath, Verily's Clinical Trial Management System (CTMS), is currently used by a wide variety of health systems to manage critical operational components of research studies.

Consortium for State and Regional Interoperability

January 27, 2023

Submitted electronically via: datacollectionforclinicaltrials@ostp.eop.gov

Grail Sipes
Assistant Director for Biomedical Regulatory Policy
Office of Science and Technology Policy
Eisenhower Executive Office Building
725 17th Street NW
Washington, D.C. 20006

RE: Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Dear Ms. Sipes,

The Consortium for State and Regional Interoperability (CSRI) is in strong support of, and eager to be involved in the initiative the Office of Science and Technology Policy (OSTP) is spearheading to enhance U.S. capacity to carry out clinical trials in emergency situations.

As stated in our response to the RFI on Clinical Research Infrastructure and Emergency Clinical Trials (see Appendix A), CSRI represents a collection of six of the nation's largest and most innovative nonprofit health information exchanges (HIEs) that share a collective vision to improve individual and population health through robust data interoperability and technology advancements. CSRI member organizations have demonstrated vast, cutting-edge capabilities to provide data-driven support to government programs and strategic priorities and solve some of the most pressing challenges associated with making clean, matched, and normalized clinical data available for research, quality improvement, and programs to improve population health.

We firmly believe that, given the mature capabilities of CSRI member organizations, clinical HIEs should be a crucial part of the "warm base" that OSTP is seeking to bolster and can serve as a core component of a data capture pilot or demonstration project. These networks were built to securely collect and maintain health data from an array of sources, including hospital EHRs, ambulatory and specialty care practices, labs, Medicaid agencies, state databases, pharmacies, etc. This existing, and near real-time source of data can support research efforts while maintaining local governance and trust.

The advantages of incorporating CSRI HIEs into an emergency clinical trials pilot are multi-fold: 1) Connecting to an HIE is an efficient way to collect data from multiple sources via a single connection; 2) CSRI members can support multiple methods of data delivery, including CDA, FHIR, flat file, HL7 v2 message forwarding and more; 3) HIEs are experts at patient matching, meaning patient medical history from an HIE has already been identity managed and collated over time offering a richer and more detailed view of a single patient than may be available from a single EHR connection; 4) CSRI's HIE members are non-profit and governed by community stakeholders. Accordingly, their data management and sharing practices have been vetted by and are subject to input from the communities in which they operate. As a separate non-profit entity, CSRI offers additional data-related services, and serves as a conduit to other state-based HIEs that are not members of CSRI given the deep relationships of CSRI members with their peer networks.

Lastly, CSRI and its members manage a significant amount of diverse population health data which is not only necessary, but pivotal in reflecting the diversity of the United States in emergency clinical trials.

We welcome the opportunity for a virtual meeting to discuss our response in greater detail and explore the role CSRI members can play in supporting and ensuring the success of this pilot. Please reach out to Morgan Honea at morgan.honea@contexture.org if you have any questions or would like additional context on the information provided.

We sincerely appreciate your consideration and look forward to speaking soon.

Cordially,

Morgan Honea, MHA
CEO, Consortium for State and Regional Interoperability
Executive Vice President, Contexture
4500 Cherry Creek S. Drive, Suite 820,
Denver, CO 80246

Appendix A

RE: Request for Information (RFI) On Clinical Research Infrastructure and Emergency Clinical Trials

Dear Ms. Sipes,

The Consortium for State and Regional Interoperability (CSRI) sincerely appreciates the opportunity to provide information on ways to build U.S. capacity to carry out emergency clinical trials and strengthen the overall U.S. clinical trial infrastructure, including potential governance models. The capacity to carry out coordinated, large-scale clinical research has been shown to be of vital importance during an outbreak of infectious disease or other public health emergency. Additionally, the need to understand the safety and efficacy of therapies within underserved populations as a way of reducing disparities and advancing equity is a moral and scientific necessity. We are eager to work alongside you to achieve our mutual goal of enhancing public health system capabilities and emergency clinical trials infrastructure to ensure and expediate the development of actionable information to address future outbreaks and emergencies in a timely, well-informed, and equitable way.

Overview: CSRI & Health Data Utilities

CSRI represents a collection of six of the nation's largest and most innovative nonprofit health data networks serving Arizona, California, Colorado, the District of Columbia, Iowa, Indiana, Maryland, West Virginia, and Nebraska. The founding members of CSRI are leading health information exchanges (HIEs) that manage the exchange of health information for over 80 million individuals, enable information exchange for more than 370 hospital facilities and thousands of healthcare providers, and are experts in data governance, privacy protection, and identity management. We believe that clinical health information sharing networks should be a crucial part of the "warm base" that the Office of Science and Technology Policy (OSTP) is seeking to strengthen and maintain.

By serving as neutral and widely trusted hubs of information, HIEs have become integral parts of the health care system. HIEs process millions of health care transactions daily and facilitate the coordination of care among an individual's multiple care providers and payors by providing the capability to electronically move health information among disparate systems. Among many other benefits, HIEs have been shown to improve the quality and safety of patient care by reducing medication and medical errors, eliminating redundant or unnecessary testing, improving public health reporting and monitoring, and reducing health related costs.

CSRI member organizations have demonstrated not only the aforementioned capabilities but have also evolved beyond these capabilities to serve as reliable data repositories that enable secure access to high-quality health data for all credentialed utility stakeholders, including states, payors, providers, vendors, and academics. CSRI member organizations have the ability to provide data-driven support to government programs and strategic priorities and solve some of the most pressing challenges associated with making clean, matched, and normalized clinical data available for research, quality improvement, and programs to improve public and population health.

Given these expanded functions, CSRI member organizations all serve as **health data utilities (HDUs)** for our respective states. While some variation exists, we serve health care providers, payors, Medicaid agencies, and public health departments. HDUs bring together health data from disparate sources

including ambulatory providers, laboratories, post-acute providers hospitals, health plans and public health. The utility cleans, matches, and attributes this data, making it available to a wide range of health care stakeholders in a given geography through standardized tools, data services and reports. Depending on a state's needs, the HDU may also serve as (or include data from) social service referral platforms, prescription drug monitoring programs, and all payer claims databases. Neutral, trusted, nonprofit HDUs serving as a public-private partnership can securely bridge and connect historic data silos to rapidly provide data and data insights to meet individual, public, and population health use cases directly aligned with the needs of OSTP.

Background: CSRI Emergency Response & Research Capabilities

Given our significant health data management capabilities, CSRI is well-suited to collect large data feeds for research purposes in a timely and secure manner. Our existing foundation of provider relationships, proven efficiency, and capacity for expansion by seamlessly linking HIEs across states offer a unique opportunity to benefit from federal infrastructure investments, while scaling quickly to meet the public health demand for novel public health emergencies. As noted in a recent blog from the HHS Office of the National Coordinator for Health IT (ONC),

> *[S]tate and local HIEs, which in aggregate receive EHR data from more than 60 percent of U.S. hospitals, could be better used as a source of patient-level electronic health data for large-scale research. HIEs routinely collect patient data from a variety of sources and then facilitate the exchange of patient health information with clinicians, public health agencies, and laboratories. Increased use of this data for patient-centered research could help facilitate research activities, including in public health emergencies such as COVID-19.*

The COVID-19 pandemic response required rapid and real-time access to transmission and vaccination data as well as bed and medical equipment availability, viewable by demographic trends, comorbidities, geography, and other key characteristics. In CSRI member states, mature HIEs served as critical aggregators and repositories for such information, enabling their states to engage in strategic, coordinated, and efficient pandemic surveillance and response efforts supported by real-time data. These networks rapidly deployed solutions including: sharing data on the spread of the virus for frontline healthcare workers; enabling public health departments to quickly gain valuable insights on trends for testing and vaccination; and, providing real-time hospital case rate and resource utilization data. Such public-private partnerships between states and HDUs not only improved the public health response but also served as important data resources for clinical systems working to treat and monitor patients.

Additionally, these clinical networks were often able to enrich data held in immunization systems by providing important contextual data – such as race, ethnicity and contact information– that have high relevance for both public health experts and policymakers. In several of our member states, state governments and public health departments have relied on CSRI data networks to populate race and ethnicity data needed for COVID-19 public health emergency priorities, such as testing and vaccination outreach, to understand the spread and response of the pandemic among different geographic and demographic populations.

While some states can leverage these existing systems, the lack of processes in place for developing emergency clinical trial protocols and for capturing trial data through consistent data elements reported across participating sites has significantly hampered U.S. capability to conduct clinical research in the face of a health-related emergency.

We believe there is an important opportunity to leverage the significant health information network infrastructure that already exists in many states to enhance and strengthen the U.S emergency clinical trials effort.

We are pleased to offer our responses to the following questions in support of this opportunity and we would appreciate the opportunity to meet and discuss these areas in additional detail.

Specific Responses

1. *Governance for Emergency Clinical Trials Response.*

The members of CSRI would stress the importance of nonprofit, state-level health data networks with existing patient and provider-level connectivity in any governance structure. The challenge of collecting data on a national level was demonstrated during COVID-19 as health providers and federal agencies alike had challenges collecting and aggregating data in real time. Efforts to rapidly scale new capabilities struggled, while many parts of the existing health care infrastructure, like HIEs, were able to stretch to meet new demands. Specifically, many HDUs have existing data aggregation, data quality, and data governance procedures supported by state and federal legislation and deployed to consumers, participants, and government agencies in near real-time. The foundational infrastructure in data sharing agreements and technology ensure a nimble response to most situations.

State-level clinical health data networks rose to the challenge of the COVID-19 pandemic, maintaining real-time detailed covid tracking databases, building vaccine reporting interfaces, and helping many small providers automate data entry to meet new reporting demands. These networks are a perfect way to maintain the “warm base” capabilities for clinical health information sharing that will be needed in a time of crisis. These networks have up-to-date networks of health providers, tried and tested technologies that have exchanged millions of patient records, and strong local and regional contacts and relationships to mitigate challenges that do arise. Many already regularly engage in support for health data research, including work on clinical trials.

In addition, local HIEs are experts in the privacy laws of their states and the concerns of their citizens, enabling the federal government to more efficiently navigate this patchwork of systems while still maintaining patient privacy and trust. Our HIEs have robust governance structures already in place including comprehensive board oversight, internal data governance, robust interoperability and quality programs, and relationships with healthcare collaboratives. These existing structures can be scaled quickly and effectively while utilizing existing relationships to maximize data sharing and trust among the health data ecosystem.

Given our health data expertise and experience addressing not only COVID-19 but also longstanding chronic disease and public health challenges, CSRI strongly requests to be included in any conversations around the development of this new national capacity and its governance. We believe it is also critical to engage HIEs at the outset of this planning process to solve for any required advanced consent or other

governance measures before any emergency actions are needed to ensure the speediest exchange of data from the HIE when needed.

2. *Identifying and Incentivizing Research Institutions and Networks; Building Diversity and Equity.*

As noted, state clinical health data networks maintain relationships with the majority of health providers in their states neutral to providers or health systems and location agnostic, meaning they are well suited to support a wide network of organizations participating in research. To understand the impacts of clinical trials, you need the widest and most diverse net of providers possible to successfully understand how small rural providers and large urban hospitals would differ in implementing the same treatments. HIEs already have these existing relationships and partnerships and thereby can assist in the outreach and recruitment of facilities for emergency clinical research studies. This state-wide presence can also bolster public awareness by leveraging the existing communication channels to help recognize and communicate any facility's commitment to the public.

Mature HIEs contain the most robust and applicable data to support accurate and rapid identification of target populations who may be needed for a clinical trial. In addition to demographic information, prior health histories, and health risk factors, many other factors could be captured from a clinical health information record leading to stronger and faster targeting and subsequent data analysis than would be captured through a siloed clinical trial effort.

3. *"Warm Base" Research.*

HDUs are perfectly positioned to support a "warm base" model of collaboration. The health data networks remain a constant and near real-time source of comprehensive health data with our CSRI members continuing to innovate to improve longitudinal health records even including social care data in some states. This existing data can continuously support "warm base" research efforts while maintaining local governance and trust. HDUs are uniquely situated to assist in identifying target infections and understanding important population differences by differentiating variables such as race, gender, or geographic location. CSRI members already support continuous research models through healthcare collaborations and academic partnerships. The involvement in a "warm base" research model to provide the most comprehensive health data for clinical research would be a direct and vital application of these robust health data networks which are a result of more than a decade of dedicated state, federal and industry investment.

Crescendo Health

January 27, 2023

Crescendo Health's Response to the OSTP RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot

For additional information on this response, please contact:

Sam Roosz
Co-Founder / CEO
Crescendo Health, Inc.
18 Bartol Street #980
San Francisco, CA 94133

sam@crescendo.health

About Crescendo Health

Crescendo Health is a public benefit corporation that provides researchers with a window into a study participant's comprehensive care journey, both historical and ongoing. Patients have a legal right to access their data held by insurance companies, health systems, and labs. Crescendo provides the tools and support to seamlessly collect patient data with participant consent and authorization. Crescendo then facilitates the contribution of this data into sponsor systems via population of an eCRF or data export that integrates with existing clinical research infrastructure. Crescendo supports sponsors in building upon their trust relationships with study participants to expand data collection in medical research, providing researchers with new tools to decipher long-term outcomes, demonstrate health economics, and improve clinical operations.

Introduction and Overarching Recommendations

The COVID-19 pandemic cast a harsh light highlighting deficiencies in our capabilities to marshal quality evidence in response to public health threats. Traditional approaches to clinical research, while powerful and reliable, are also too often slow, inefficient, and fail to generate evidence that reflects the vast diversity of the United States. We applaud the efforts of OSTP to learn from the lessons of the past several years and invest in infrastructure that strengthens the US's ability to respond effectively to emergent public health threats while also accelerating American medical innovation.

The 21st Century Cures Act and recent rules from ONC and CMS have afforded a valuable opportunity to implement new models of clinical research that leverage data already being

generated in routine clinical care for research. Of particular interest are recent advancements a patient's right to access their own health data at locations across the United States. Patient access APIs mandated for many health plans and all certified EMRs through CMS and ONC rulemaking means that for the first time it is feasible to efficiently assemble a comprehensive and longitudinal health journey for consenting trial participants. Furthermore, the Office for Civil Right's enforcement actions against covered entities engaged in information blocking means that phone/fax-based methods of data collection are truly reliable to cover any gaps APIs can't fill. This opens up the possibility of a new research model where as patients consent to enroll in a trial, they also sign appropriate forms and activate APIs that facilitate data collection for both their use and for research.

This patient-centric approach to evidence generation offers many advantages for US resiliency to public health threats:

- As this approach can be conducted in a fashion where data is collected and parsed at a central hub, it can be quickly deployed in collaboration with a wide range of enrolling sites with de minimis infrastructure and staffing needs. This means that in addition to a "warm base" of active sites, new sites can be quickly activated based on the unique characteristics of the emergent threat (e.g. regional density or specialist participation).
- Where appropriate, this method can also support a direct-to-patient recruitment approach for a decentralized study that doesn't require any participating brick-and-mortar sites
- Through this model, the data available about a study participant isn't limited to just the site that enrolls the patient, but rather the full breadth of the patient's encounters with the US health system. Payor claims offer an opportunity to detect potential signals of interest that occur beyond the trial site, and based on those signals additional data can be collected to support confident outcomes assessments.
- Data collected in this model is available in a centralized hub, with specific elements disclosed to researchers based on their needs. As learning advances, researchers can "go back to the well" for additional data on the enrolled cohort to test new hypotheses without delay.

We suggest a pilot that incorporates these patient-centric methods of data capture in order to maximize the flexibility and robustness of the US's emergency clinical trials capabilities.

Responses to Select Questions Posed in RFI

2. HL7 FHIR APIs. We seek comment on how U.S. Government and external stakeholders might leverage FHIR APIs to support research in emergency settings as well as in the pre-emergency phase, and in what areas further advances might be needed.

As discussed in the introduction above, patient access FHIR APIs leveraging OAuth 2.0 authentication offer a path to efficiently capture clinical data from a wide range of care settings in a fashion that can be quickly deployed across a large population. The APIs in their current form, however, are not able to collect health data from all institutions. While CMS rulemaking has led federally-funded exchange plans, Medicare plans, Medicaid plans, and some large

commercial plans to implement APIs for patient access, some large payors and many regional commercial health plans have not yet followed suit. Similarly while USCDI standards have brought structure to some EMR datasets, many potentially valuable data are still either inaccessible due to lack of health system implementation or data feeds that lack metadata context to enable effective parsing. Another complicating factor is variance in the duration that access tokens obtained during OAuth 2.0 authentication remain viable for querying data holders. Some expire shortly after initial queries, while others remain evergreen.

These hurdles can be overcome at additional time and expense through additional data partnerships, fax/phone-based data collection, and recontacting participants. In order to make this approach more robust and efficient, we suggest that regulators work with industry to expand the reach of patient access APIs, ensure metadata is available to support downstream parsing of data, and ensure that access token viability after patient consent is extended (particularly for emergency and research contexts).

6. Consent, deidentification, return of results. The use case in this RFI contemplates that data would be managed through a central repository or repositories and made available to researchers beyond a patient's home institution.
b. We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information. It would be helpful to receive comments on which deidentification and protection approaches are sufficiently mature to support a pilot effort in the near term.

It's important here to note some of the inherent limitations of "de-identified" data sets. De-identification is a term of art under HIPAA that can be facilitated through two approaches:

- Safe Harbor: This approach involves the removal of 18 data elements, including elements often critical for effective public health research such as any date information more specific than the year (creating meaningful challenges for assessing ordinality or tracking progression of a public health threat)
- Expert Determination: This approach requires an expert statistician assert that the risk of a patient being re-identified in the context of data use would be "very small". In practice, this approach limits available de-identified data to just those fields which are structured. Free text notes like those found in clinical notes, pathology reports etc are typically removed entirely for any de-identification at scale and these notes can be critical for understanding emerging threats where appropriate structured fields don't yet exist.

Instead we would propose a focus on "limited data sets", which replace patient identifiers with subject IDs but are still considered protected health information under law. Aggregating data in this form mirrors the approach taken for the vast majority of prospective research studies and ensures that the richness of data is preserved to support necessary queries while still protecting participant privacy.

d. We seek comment on any technical capabilities that could support return of results to study sites or participants, where appropriate.

If the patient-centric methods described above are deployed, this also offers a "hub" that maintains the identities of participants and can be utilized to recontact patients to invite them to participate in additional studies (if the initial consent supports such activity) or return results to the participants.

7. User interface and experience. With all of the above technologies, we seek input on: b. To the extent a particular form, app or other tool requires input from a health care provider or other user, the best ways to increase the likelihood that users will actually provide that input. It would be helpful to receive comments on methods that are available for completing empty fields after the fact, or otherwise managing any missing data.

For patient-centric approaches, incentives / payments can be considered as a mechanism to support yield (with appropriate case-by-case ethical review).

One meaningful advantage of API-based data collection is that a given access token can support both retrospective as well as prospective data collection. This means that many data can be collected without ongoing active patient engagement and any retrospective gaps can be filled at the next patient touchpoint if and when it occurs.

We suggest a focus on obtaining claims data from a payor as a starting point, which provides insight into the full patient journey during a coverage period and can be used to identify possible gaps in collected clinical data. Where API connections are unavailable / broken / inactive additional methods of data collection such as data partnerships and phone/fax-based methods can be used to address any residual critical gaps.

12. Specific commercial capabilities. Commenters who are developing a technology or product that might be relevant to any of the topics set forth above are welcome to include a description of that product. Comments about a specific technology or product should be limited to three pages or less.

Crescendo Health offers a set of tools that helps researchers obtain comprehensive and longitudinal outcomes data from the full patient journey for consenting trial participants. Our approach is designed specifically for clinical research and is built by an experienced team of clinical trialists and software leaders. Crescendo's approach is protocol-driven, meaning that our implementation can adapt to the unique needs of a given study (e.g. supporting either site-based consent process or a Crescendo-hosted eConsent process).

Once consent is obtained (either through Crescendo or at a study site), patients are directed to complete a 5-10 minute online onboarding process to sign appropriate HIPAA forms and activate available API connections. Crescendo then facilitates the collection of the data for the participant's own personal health record, and automatically makes the data available to the sponsor if supported by the informed consent. Crescendo uses multiple methods to ensure longitudinality and completeness of resulting data, including APIs, data partnerships, and phone/fax-based data collection. Crescendo parses available data and can populate an eCRF

(or output properly formatted data for bulk import) for sponsor use, including removing personally identifiable information and replacing it with subject IDs. As directed by researchers, Crescendo can also recontact patients to return results or communicate a separate study opportunity.

Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Federal Register: Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Thank you for allowing the Keyrus Life Science USA (KLS) and Keyrus USA, Inc. (MAS: 47QTCA22D006B), known collectively as Keyrus, to respond to this RFI, 87 FR 65259. We are very excited to be a part of this RFI and thank you in advance for the opportunity to partner together.

I would like to begin by introducing myself to your team. My name is Karen Marie Josey and I am the Senior Director, Business Development at KLS. I will be your main contact moving forward. I have 30+ years in the commercial and clinical pharmaceutical world. As we look at the exciting possibility of partnering together with OSTP, AP3, ONC and NSC to establish a U.S. level governance structure and outreach to a wide range of institutions, clinical trial networks, and other potential trial sites that can participate in emergency research, both domestically and internationally, I will use my industry experience, contacts and connections to bring the most innovative and effective solutions to you and your teams. I will be joined by members of the KLS clinical trial team and our Keyrus USA data specialists to support the expansion of clinical research into underserved communities, and increase diversity among both trial participants and clinical trial investigators. Building U.S. capacity to carry out emergency clinical trials will enlarge and strengthen the U.S. clinical trials infrastructure overall.

Keyrus Life Science is a globally connected CRO bringing life data sciences and digital enablement together to fully leverage the clinical research ecosystem and real-world evidence in healthcare, making clinical research activities more reliable, innovative and agile. We provide a full suite of clinical trial services to optimize patient recruitment and engagement, to leverage insights from data, and to unlock new horizons for personalized therapeutic approaches. At Keyrus we knack for ethical innovation. The Keyrus Innovation Factory is our innovation incubator that operates on an international scale. We iterate on the latest use cases and technological trends with a spirit of respect, fairness, and progress. Empowering clinical research with data to answers biggest health challenges. Data is the key driver of innovation and the foundation for solutions that enable care improvement and augment clinical research capabilities. We use our connected data approach to promote faster translation of R&D efforts to patients, to improve personalized, predictive, preventive and participative approaches to health concerns.

With more than 25 years of experience, Keyrus makes data matter to address the biggest clinical challenges in a positive way to enable long-term success. The graph below is a sampling of the support and management Keyrus can provide to OSTP and all agencies, sponsors, sites and patients involved and effected by an emergency clinical trial.

Keyrus Life Sciences

Unique Positioning to Support OSTP Data Collection for Emergency Clinical Trials and Interoperability Pilot

Frame
Clinical Trial Strategy

- Protocol design & development
- Feasibility studies
- Regulatory strategy
- Pilot development and execution

Operationalize
Clinical Development

- Regulatory affairs
- Clinical operations
- Biometry
- Medical writing
- Pharmaco-vigilance

Implement
RWE and Late Phase

- Observational studies
- Secondary data use & external data intelligence
- Compassionate and early access program

Enable
Life Data Sciences

- Clinical data fabric
- Artificial intelligence
- Data visualization
- Sensors & wearables
- Automation

Reach
Digital Enablement

- Hybrid & DCT
- State-of-the Art simple Patient engagement
- User friendly app engaged to allow diversity and inclusion in the patient sampling

Like OSTP and all the U.S. affiliated agencies focused on clinical trial emergency readiness, at Keyrus we not only believe it is the data itself that matters, but the problems we can solve by leveraging it. By “making data matter,” we don’t exclusively mean in a clinical trial performance context; we make your data matter from a broader, human-oriented perspective that enables positive change on a larger scale. The kind of broad scale perspective needed in the event of an emergency healthcare crisis in this country.

At Keyrus, we plan to use our scientific and technical expertise to empower OSTP with actionable data-driven insights. Beyond simply understanding data, we use it as a driving force for progress and innovation - a means to a better and healthier future. As data plays an ever-expanding role in all of our lives and across clinical research our experienced team can be there to help OSTP design the interoperability pilot, execute it and analyze the results after the conclusion of the pilot. To us, data is a window into our world, its workings, and the way humans interact with and shape it. Data is the story of our past and the script for our future, making it inherently human. This approach allows our clients to put more focus on the individuals they serve. More broadly, it enables them to use data in a way that will positively shape the future. This is why we focus on extracting insights and value from data - we know it has the ability to move us forward in a positive direction, not just economically, but environmentally, socially and across the most pressing health-related challenges.

If given the opportunity, we would approach your interoperability pilot with both present and future-oriented. We would implement solutions that solve current OSTP challenges and add immediate value while also looking ahead at future opportunities for innovation and progress, with a focus on emergency conditions.

This would enable OSTP to proactively reinvent your clinical trial strategies, to the final benefit of the patients and the country while working in a timely manner. We believe data is the raw

material that OSTP will need in order to stay ahead of current and future health crisis occurrences. We are experts at tackling complex problems and providing our public health clients with straightforward, effective and scalable solutions.

Keyrus has always had a focus on diversity and inclusion. We believe a better future begins when we bring the best together. This means a constant emphasis on diversity and inclusion in and outside of the workplace, and persistent dedication to continuing to learn and improve which we would bring to OSTP and all related agencies we would partner with during Emergency Clinical Trials and/or the Interoperability Pilot. Putting our beliefs into action, Keyrus has made a point to develop a strong not for profit (NFP) data practice, providing steeply discounted services to help modernize NFP's infrastructure and reporting capabilities. This enables NFPs to direct grant/donor money to the impact of their mission to provide services to underserved communities.

Keyrus has worked with other governmental agencies that were preparing their systems in the event of emergencies, in these situations, it is critical that these systems are regularly tested and go through real-world simulations to ensure operability in a true emergency scenario. The solution itself needs to provide flexibility for quick configuration changes because these types of emergency scenarios without fail will provide variables unaccounted for. Out-of-the-box solutions typically will be quicker to stand-up but will break much easier when they encounter unexpected behavior or functionality that was not planned for.

One of the other main features of our solution is that we would be providing a completely codified platform that follows best practices in regard to development operations (DevOps). The benefit of leveraging these techniques is that the entirety of the platform (infrastructure, software, and configuration) can be activated programmatically removing human error and vastly increasing the speed and reliability of spinning up the platform when needed. We are happy to give you a capabilities presentation and demo of this technology so you can see the benefits for the program and solution you are building for emergency readiness.

Below you will find the Keyrus solution responses to areas of need for OSTP and related agencies for Data Collection for Emergency Clinical Trials and Interoperability Pilot. We would be very happy to engage in a meeting where our CRO and data team can expand on our ability to support the OSTP team and be 'ready together' for any health crisis that may arise.

1. A U.S.-level governing entity would oversee development of a clinical trial protocol for broad distribution across clinical trial networks and sites.

Keyrus Life Science, is a connected global CRO with 30+ years of Phase I-IV clinical trial experience. We work with clients from the beginning discovery stages all the way through approval and multi-indication approval. KLS and our subject matter experts have vast experience in overseeing the development of all clinical trial protocols, qualifying our large network of clinical trials sites, training and ready preparing those sites and staff and monitoring the sites and staff throughout the entire clinical trial study process.

We have contracts in place with over 20 large site network groups in North America and can identify, train and engage those site groups within 3-6 weeks. We continue the education and evaluation process during the clinical trial and have a state-of-the-art technology platform that can oversee and evaluated the site staffs understanding, ability and readiness as it relates to the OSTP protocol.

Our site 'Ready' approach is a more effective and efficient way to accelerate enrollment and minimize risks. It helps prevent issues downstream by delivering better training that predicts and improves site and study team performance. This innovative technology streamlines site initiation and reveals which teams and sites are best prepared to successfully conduct a study. In the event of an emergency healthcare crisis in the U.S., both time and preparedness will be key factors.
We employ a behavioral science-based approach that enhances role specific training and improves performance. When you combine our massive and ready site network, with the quality control 'ready' approach, OSTP can confidently move forward with an emergency clinical trial initiative knowing that KLS will manage the site setup and training in an effective, efficient, and timely manner.

2. HL7 FHIR APIs

In all use-cases that require a disparate group of users for the collection and analysis of data, there needs to be data structures and formats that all parties adhere to. This is precisely what the FHIR framework accomplishes to make the sharing and accessing of data possible across different sites, user groups, and electronic health records.

The platform would be built leveraging AWS API Gateways to process RESTful APIs that adhere to FHIR requirements to collect data on the different data objects that we can expect to store (E.G. patient, observations, organizations, etc.). The resource definitions from FHIR would map to their own API Gateways to ensure the expected structure of the request is coming through. Using the api managed service by AWS enables by default 10,000 requests per second, with the fine-grain control of how we throttle the number of requests (data) based on what our downstream architecture requires. This enables the platform to handle high peaks of traffic in an emergency situation with full reliability. Also, in tandem with other services like AWS CloudTrail, OSTP will maintain full visibility into every request that traverses through these endpoints. We would also apply rules and security measures to notify and protect against unwanted traffic / threat detection (we would leverage Guard Duty which is AWS's service that uses anomaly detection and ML to identify threats).

This platform supports bulk access to electronic health records by providing the flexibility of features we can enable into the API and subsequent services since we are building this in a custom manner. The advantage of using serverless managed services from AWS is that the resources will autoscale to ensure performance to a wide range of traffic patterns (e.g., emergency utilization vs. pre-emergency utilization). Interactivity with EHR systems will be possible due to the FHIR layer being the bridge between source and target. But investigation on what EHR systems OSTP would want to leverage would need to be done.

Are these systems that are locally administered from the different organizations and sites that sit behind their own firewalls/network? If so, the effort would be high in working together to coordinate security rules for each and every target and would recommend visiting the standardization process that all members of this initiative use to house and share data.

Another recommendation for accessing the data and general analytics would be to host AWS QuickSight dashboards that sit on the centralized deidentified data sets. For the end-users that would access these dashboards, they would typically go through an application that has the reports embedded within it to ensure network/authentication security.

3. Clinical trial data is typically sent to the trial sponsor though an electronic case report form (eCRF), which is the record of data that is required under the protocol to be captured for each trial participant. A data element in an eCRF is the smallest unit of observation for a particular subject

Our eCRF system is designed with its end Users 'needs in mind. Our System not only allow for data collection, but is fast and Easy to learn exactly what we need in the event of a crisis. The system, supports both MedDRA and WHODrug dictionaries, provide flexibility to handle medical device and drug logistics. Both device and drug logistic workflows are connected through our EDC (Electronic Data Capture) and RTSM (Randomization Trial Supply Management) systems. The system is certified CDISC and uses the most advanced web survey software allowing you to create simple or complex follow-up questionnaire designs (ePRO, eCOA). Build online surveys, forms, polls quizzes, questionnaires or use existing study forms to be filled out by patients via email or SMS, using a tablet or a smartphone. Our system includes a fully compliant eConsent solution that allows for both, in- person digital consenting and remote consenting of patients. The eCRF system is Compliant with FDA 21 CFR Part 11 and GCP.

4. The eCRFs would be transmitted electronically via common APIs to the sponsor

Seamlessly connect operations with data management. Get simple access to a suite of applications, while open architecture ensures rapid integration with third party applications via API. Leverage connected sensor technologies including Watches, Scales, BP monitors, Pulse Oximeters, Sleep Mats and more enable sites to remotely monitor patients and make data-driven decisions in real-time.

5. The study site's health IT system would present the eCRF content to clinicians in a manner that expedites data collection and (ideally) fits within clinician workflow

We routinely expedite data collection through and to clinicals, sponsors, sites and agencies. We do this within the constraints of the clinician workflow. Working with OSTP, the sites, the sponsors, and our Keyrus team we would have a dedicated team in place to monitor all eCRF content to make sure it is collected in real-time accessible to all parties.

The KLS eCRF system provide a flexible software that enables easy study set-up and management. Customers and end users value the simplicity of customizable workflows.

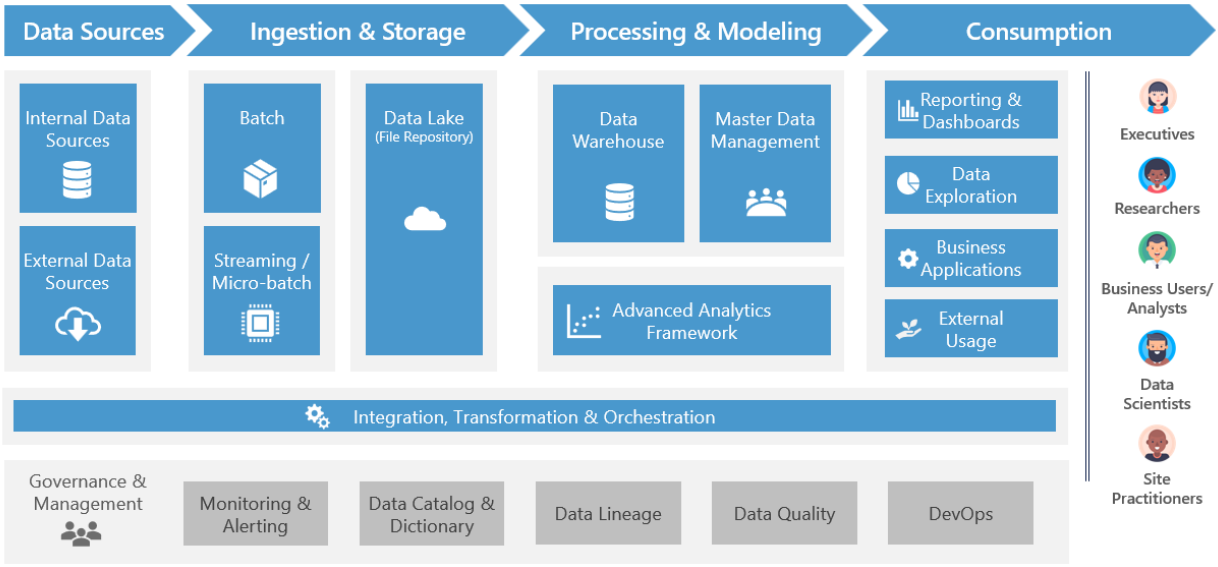
According to our clients, one of the key advantages of using our system compared to other EDC tools, is its quick implementation. A study set-up is possible in weeks not months.

6. Consent, deidentification, return of results

The KLS eConsent software simplifies the consent process, raises patient comprehension and retention, eases workloads for study teams and sites. Flexible and powerful design features allow for the creation of sophisticated and intuitive electronic informed consent forms (eICF) with videos, graphics, and downloadable PDF. The necessary data (both structured and unstructured) will flow securely downstream from the eConsent software into the central data platform.

Informed consents and/or authorizations would be stored in their own document store within the data lake layer (S3). The data regarding that document (patient) will be passed along and used as authorization as we pass data further along into the more centralized layers. The ability to integrate the data from this layer with downstream rules ensure that no patient records without completed consent forms gets centralized together.

The graphic below shows the layers of a high-level modernized data analytics platform. The solution being proposed will be built out natively from AWS's services (a design of the AWS services being leveraged can be found in Appendix) providing a high level of integration between each component.



The downstream central repository would be AWS's modernized data warehouse service Redshift. Similar to the other services recommended (and cloud in general) the advantage of Redshift is the ability to automatically scale up the storage and compute needed to provide consistent performance for the queries and analytics of the data consumer.

The platform would have different data layers in which transformation, de-identification, and centralization of the data would occur.

Each of these layers would apply encryption of the data at rest (S3 and Redshift provide this functionality as a standard capability). Data would be encrypted in transit using TLS/SSL. Following HIPPA's "Safe Harbor" provisioning, services like AWS Glue and Lambda would de-identify PII data as the data becomes more centralized and available for consumption. Depending on how the data asset is classified dictates the type of deidentification method (masking, hashing, encrypting, tokenizing). The platform would also leverage AWS Macie to continually crawl and evaluate data assets and notify potential unmasked data assets across the different services to ensure real-time security/monitoring.

Within Redshift, additional transformations would be done to create an integrated and usable data model for analysis across studies and organizations as well as serving other types of data objects like aggregated reporting tables and materialized views. The AWS Data Catalog will be ingrained in the AWS Glue process which will provide an automated way to monitor your data assets and data definitions. The platform would rely on mapping the EHRs back to the patient authorization to enable that data to be deidentified yet still used to create benchmarks and aggregate measures across all of the different sites.

Lastly, using a front-end technology to create visual representations of the data to provide results across patients and sites will be needed. QuickSight is the AWS service for this. QuickSight will also scale to support peak demand and is easily embedded into front-end applications that could also be hosted with AWS. QuickSight keeps this as a full native solution but other market leaders in this space like Tableau could be used as an alternative to this technology based on end-user preference/needs.

7. User interface and experience

At Keyrus, we believe that engaged patients mean a better trial experience and better trial results. KLS can provide for your trials in an emergency readiness position with the next generation eCOA/ePRO and engagement platform designed to engage the clinical trial site staff and the patient and their caregivers from the first moment of contact.

In past clinical trials when KLS employed this next generation system, study participants had >90% compliance. This Compelling onboarding engaged participants to continue in the study with only a 1% attrition rate. High engagement, low attrition and faster data capture will improve quality of results and drive new insights for OSTP, most especially when an emergency health care crisis arises, means the difference between getting you need to save lives. Keyrus is committed to bringing OSTP the most innovative technology and tools to design, execute and analyse clinical trials, in a timely manner from Study start up thru trial, completion and drug approval.

Because of our vast site group partnerships and established contracts and our one-stop eConsent platform that can be used and viewed by patients, sites, staff, sponsors and OSTP in real-time, we know from experience that this approach is the best ways to increase the likelihood that users will actually provide the input needed for efficient and effective data capture. Keyrus is very excited to be able to bring our teams together to demonstrate this next generation technology cloud-based ePRO/eCOA platform, APIs and workflow integrations

support which we have used to deliver large-scale clinical trials across the globe. Integrating with other clinical trial technology and services, we can bring the high rates of data capture compliance and participant retention we have seen with other partners to OSTP for your current and future project.

8. Capturing data elements required for clinical trial protocols

Our eCRF is a s an advanced electronic data capture (EDC) and clinical data management system (CDMS) for capturing, managing and reporting clinical research data optimizing data collection by efficiently streamlining clinical data collection process. The software has been designed to meet the needs of CROs and pharmaceutical companies, government agencies as well as research institutions and academic researcher. Our system allows for eDiary, ePRO, eCOA, and eConsent which are mandatory during a crisis. The system also allow extra data capture via an open Architecture And Easy Integration with third party applications via API.

Our recommendation is to build the ingestion framework (the part of the solution that will capture the data from operational assets like the FHIR Questionnaire or a SMART form) using AWS managed services. Whether it's for ingestion, transformation, or load, we can build our solutions to use the most generic solution which will create an emergency/research system that is flexibility and is easy to maintain/operate. These generic solutions require configuration files, tables and parameters that control widely used rules and logic so in the event we need to add new functionality, we update it in one place, and those rules trickle down to all of the relevant services. For example, as we build out the API gateways and Lambda functions that bridge and/or validate the structure of the API request as well as its content (E.G. FHIR Questionnaire), the overlaying logic and rules will be managed in one place. The services that need to then have more custom logic will be stored and hosted at the individual service instead of leveraging the upstream configuration files.

Using a tool like a FHIR Questionnaire, FHIR Questionnaire Response, or SMART form is a great way to capture the required data elements. The more you can standardize the approach to capture data across the different actors, the better. This leads to least custom code across the different methodologies which will lead to better administration of the platform. With these types of implementations, we will need to create data validation rules to ensure the data is coming in with the necessary structure. These rules also check the values of these responses to ensure data matches expected responses or falls within certain thresholds.

9. TEFCA and QHINs

As noted above, TEFCA is in the implementation phase at this time. In the future, the TEFCA QHINs are expected to support implementation of the FHIR APIs (see the ONC Recognized Coordinating Entity's January 2022 FHIR Roadmap for TEFCA Exchange [6]). We would appreciate comment on the opportunities and challenges regarding development of API implementations toward the use case described above, particularly given the current status of TEFCA and QHIN participation. Specific topics in this connection include the following:

a. Certain policy and/or technical constraints will need to be specified for currently authorized Exchange Purposes under the Common Agreement (e.g., Public Health). We seek comment on which of these constraints will also be applicable to a future research-focused Exchange Purpose.

b. Opportunities that may exist for using the initially authorized Exchange Purposes to accomplish the use case described in this RFI.

c. How the Public Health Exchange Purpose could be used to advance the goals of this RFI; what aspects of the use case described above might fall within the scope of the Public Health Exchange Purpose.

d. How a future research-focused Exchange Purpose could be structured to advance the goals of this RFI.

e. Other opportunities or constraints related to TEFCA that should be considered with regard to this RFI.

10. Emerging technologies

The data infrastructure that Keyrus is presenting in this RFI is consistent with standard data practices for data management, regardless of the technology. The AWS toolset that enables data ingestion, transformation, storage, and reporting can be replaced by 'best in breed' or industry disrupting tools that emerge later. The key is the data is stored in a standard relational schema so it may be migrated to other data storage that becomes available, or accessed by authorized tools for analysis without additional manipulation of that data.

With regard to the storage methodology, in our experience when assessing whether to adopt new types of data management technologies, one must weigh the following considerations.

- Availability of skills in the market
- Longevity of the technology

These two risks lead us to recommend an agnostic format such as a relational database for this initiative which will enable a seamless migration in the event of moving to a different technology in the future. By following standard data warehousing practices, OSTP can ensure that the flow of data is reliable and timely in the case of an emergency clinical trials situation. In the case of an emergency, the government will be able to enlist support from any number of proven Data/IT Services suppliers as the methodology is ubiquitous across the Data industry. For example, technologies like Apache Hadoop or MongoDB deviate from the standard relational data model, and require specialized skills to manage; limiting OSTP's pool of resources in an emergency situation.

The collection of data will certainly be impacted by future technologies, whether they are more advanced hand-held devices, or smart monitoring devices deployed in the field. The infrastructure which Keyrus is recommending for the data management aspect of this solution will easily adapt and scale to accommodate new data as needs/capabilities evolve in the clinical trials – see our response to question 8 with regard to configuration tables.

11. Pilot or demonstration project

i. A demonstration project with commercial partnership

Keyrus brings expertise in the field of Clinical Trials, as well as Data Engineering from deploying solutions across the private sector. This POC is an opportunity to partner with a commercial firm that offers a unique perspective of having both Clinical and Data experience that are incredibly pertinent to this solution.

ii. A public-private partnership

Keyrus has worked globally with public health teams in all therapeutic areas and all stages of clinical trial readiness. During the pandemic, KLS, ran a COVID related trial and was audited by Health Canada during the study. It was the first virtual audit Health Canada or KLS had ever been a part of. KLS came through the audit with no findings and was highly praised by health Canada. We have global contracting experience with public agencies and we have been accoladed at having a main stream, flexible and relevant contracting process. Whether it is a LOI, consulting contract, MSA or full study contract, we will be able to execute any documents needed and react quickly to any emergency study needs.

iii. An agency-funded program

We have experience in working with agency-funded partners, clients, patient advocacy groups and disease centered non-profits agencies. We have written for clients and partners grant requests and have complied and with global rules and regulation for agency-funded programs.

Thank you for the opportunity to share our innovative technology solutions and experience in clinical trial readiness as it relates to the OSTP need for a data collection plan for emergency clinical trials and an interoperability pilot. We are ready to support the efforts of OSTP and related agencies so that a comprehensive and real-time plan for execution is ready if and when it may be needed. We are excited about the opportunity to meet with your team to further discuss the capabilities of Keyrus Life Science and our potential partnership for the future.

For further discussions please contact:

Karen Marie Josey
Keyrus Life Science
Senior Director, Business Development
810-610-4806

www.keyruslifescience.com

HealthEx | healthex.io
80 S Park, San Francisco, CA

January 27, 2023

Submitted electronically

Office of Science and Technology Policy
The White House

Re: Request for Information (RFI) on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Dear OSTP & ONC,

Thank you for the opportunity to provide comments on the OSTP's RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot.

Briefly, HealthEx is a startup company, in the healthcare IT space, founded in Spring 2022 with the support of General Catalyst and Electric Capital. At HealthEx, we are building a novel health data collaboration ecosystem that is designed to be an ethical and transparent global good, with realigned incentives that better serve health systems, patients, and researchers in support of precision-medicine discoveries. We support efforts in the clinical trial and observational research space, and HealthEx aims to be the health-system facing, patient-centric data collaboration ecosystem that serves as an efficient front-door for researchers to access multi-health system clinical data of the highest quality.

Core to HealthEx's mission and values are engaging patients as core stakeholders in the research and clinical trial space.

Based on HealthEx's deep efforts in the research and clinical trial space, below is our feedback and comment to achieve the goals outlined in the RFI on Data Collection for Emergency Clinical Trials and Interoperability Pilot.

Focus area 1: How to optimize data collection for clinical trials carried out across a range of institutions and sites, both in emergency settings and in the pre-emergency phase.

A key challenge in the clinical trial process is the identification of patients to enroll and participate in clinical trials, particularly patients from underrepresented demographic groups, including women and other minority groups. A broad and representative patient population enrolled in a clinical trial is key in optimizing trial result applicability to a broad patient population.

Supporting the identification of patients who may be participants in clinical trials is a known challenge, with the vast majority of trials under-enrolling patients from various minority groups; for example a 2018 JAMA Oncology study noted that black patients made up only 3.6% of participants in clinical trials for oncology drug approvals during 2013-2018, in stark contrast to black patients' proportion of US cancer incidence (noted to be 22% of total).[1]

A mechanism for identifying eligible patients across clinical trial sites in a manner that is secure, trusted, and that requires minimal egress of data from health systems will be beneficial in supporting the prompt identification of patients who are eligible to participate in trials. Ensuring that the underlying data access mechanism follows regulatory standards of what data should be available in what format for each patient will allow for more predictable outcomes and analysis. In addition, Patient Reported Outcomes (PROs) are gaining popularity and may offer non-traditional health data that are not usually available via Clinical data from an electronic health record (EHR), and involving patients more actively in the trial process via novel engagement mechanisms will be beneficial to the data quality of trials, overall.

Focus area 2: OSTP and ONC are seeking input on viable technical strategies to distribute clinical trial protocols and capture clinical trial data using common application programming interfaces (APIs).

Traditionally, clinical trials require principal investigators (PIs) to identify patients fulfilling detailed clinical criteria to determine who may enroll in a trial. The detailed clinical codes (condition, lab values, etc.) and standards involved in being able to construct such queries electronically require deep technical knowledge, which is out of reach for most

[1] Loree JM, Anand S, Dasari A, et al. Disparity of race reporting and representation in clinical trials leading to cancer drug approvals from 2008 to 2018. JAMA Oncol 2019;5(10):e191870-e191870.

PIs and clinical institutions looking to enroll patients in trials. As a result, much of clinical trial enrollment today is a manual process for individual PIs enrolling patients in studies, or for study personnel to work with individual PIs to manually identify eligible patients. Once a patient is enrolled, the clinical information required for a trial is typically then entered by study personnel.

HealthEx is building a superior user experience for both PIs and patients looking to enroll and participate in trials, through the use of modern APIs available, including:

- Bulk FHIR API for accessing large population level cohort data: (https://hl7.org/fhir/uv/bulkdata/). This approach is applicable to large populations.
- FHIR APIs (USCDI-compliant) served by Health System EHR FHIR servers to fetch individual patient level data in a manner that is health system initiated. This approach is applicable to single patient use cases.
- SMART on FHIR APIs (https://hl7.org/fhir/smart-app-launch/) to support patient-initiated authentication and authorization for automatic retrieval of patient demographic information, clinical diagnoses, lab values, and other clinical information relevant for patients interested or enrolled in a specific trial.
- Claims data with ICD-10 codes via beneficiary authentication from APIs such as the CMS Blue Button API (for Medicare) or 3rd party services such as Flexpa.
- Unified Medical Language System (UMLS) Crosswalk REST API to map Clinical/SNOMED/ICD-10 codes to Clinical Trial Search API-compatible NCIT disease codes that can be used to query the Clinical Trial database

All of these APIs offer cost-effective, efficient approaches to obtain patient data from their local points-of-care, in order to identify eligible patients for trials, and once enrolled, to share data to the study sponsor.

Focus area 3: OSTP and ONC also seek information about whether there is value in a pilot or demonstration project to operationalize data capture in the near term, for example within 6-12 months of the close of comments on the RFI.

HealthEx welcomes the opportunity to participate in a demonstration project that operationalizes data capture in the near-term. We are partnering closely with leading health systems and not-for-profit groups to support patient identification and multi-health system data collaboration in a federated manner that is secure, privacy-preserving and auditable. At present, we are designing demonstration research efforts that highlight

how the HealthEx data collaboration ecosystem can be used to increase diversity in clinical trials, in partnership with a nationally recognized cancer center and a nationally recognized cancer-fighting not-for-profit organization. HealthEx welcomes the opportunity to share details of this effort with the OSTP and ONC.

As we have learned from the COVID-19 pandemic, timeliness of clinical trials is incredibly important to ensuring future health and safety of individuals. Data collection and data collaboration are, correspondingly, foundational to timely clinical trial efforts. We appreciate the opportunity to provide comments on this RFI, and we look forward to working with the OSTP, ONC, and others on this important topic.

Sincerely,

Priyanka Agarwal, MD, MBA
Co-founder & CEO, HealthEx
priyanka@healthex.io | **in**

Anand Raghavan
Co-founder & CTO, HealthEx
anand@healthex.io | **in**

Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot: Using OneSource to Enable Rapid Deployment of Clinical Trial Data Capture

Prepared by:

Quantum Leap Healthcare Collaborative

OpenClinica, LLC

In response to:

RFI: Data Collection for Emergency Clinical Trials and Interoperability Pilot

Doc. 2022–23489

I. Introduction

In detecting and in response to public health emergencies, the timely availability of quality data is paramount, as is the ability to share and analyze data pooled from multiple clinical sites and in the context of the current state-of-the-art in relevant therapeutics. Manual data capture, cleaning and validation in clinical trials entails a significant investment of time and resources, the need for source data verification (SDV) causes delays in the availability of quality data, and complex study startup requirements delay launching of trials. Automated data collection can help alleviate these problems, but current lack of interoperability in health data systems and the complex nature of research data impedes the sharing and reuse of data across multiple nodes.

With support of the Food and Drug Administration (FDA) and the Biomedical Advanced Research and Development Authority (BARDA), Quantum Leap Healthcare Collaborative (QLHC) and OpenClinica deployed the OneSource program to streamline and automate the collection and reuse of clinical and research data at the point-of-care. OneSource, based on the OpenClinica Unite technology, provides an electronic health record (EHR) to electronic data capture (EDC) integration solution that captures regulatory-grade data from SMART on FHIR compliant EHRs with both provider- and patient-mediated data access. OneSource is deployed at 15 large hospitals and health systems in the U.S. supporting multiple trials, including the I-SPY COVID Trial, which seeks to save the lives of critically ill COVID-19 patients, and has led to 60% time savings on data entry tasks, significant reductions in error rates, and streamlined workflows for users. OneSource utilizes FHIR-based US Core Data Interoperability (USCDI) standards to permit structured data capture directly from the EHR for reuse in clinical trials, registries or other secondary uses.

By establishing the capture of USCDI data elements as a standard part of clinical care between public health emergencies, OneSource, and site networks based on similar standards, enable automated capture of regulatory grade data for clinical trials, including patient characteristics, diagnostics and diagnoses, laboratory measures and outcomes, at all clinical sites. Direct capture from the EHR without human intervention significantly reduces human and financial resources required for data capture, while vastly increasing data fidelity and reducing the need for source data verification. This means faster access to better quality, highly portable data. Furthermore, in times where no emergency exists, OneSource can facilitate the creation of low-cost, low-maintenance registries that can be used for monitoring/surveillance purposes, to detect outbreaks on a national or local scale, and establish

baseline outcomes that may provide important guidance on potentially effective treatment strategies early in the response to the emergency.

OpenClinica was founded in 2006 as an Electronic Data Capture (EDC) platform and has supported over 10,000 studies, including clinical trials on six of the seven continents spanning Phase I-Phase IV research. The technology is a cloud-based, modern EDC along with modules supporting reporting, ePRO, Randomization, EHR to EDC integration, and patient-directed health record sharing.

Quantum Leap Healthcare Collaborative (QLHC) is a 501(c)(3) charitable foundation supporting the development and implementation of innovative ways to deliver better, less costly healthcare. QLHC has successfully established unique partnerships across the medical, technology and bioscience industries, as well as the federal government, all necessary components to accelerate healthcare research into the marketplace. QLHC's efforts focus on quality-of-care and quality-of-life issues and creating initiatives that foster excellent clinical practices using quality improvement disciplines with a strong patient-centric focus. QLHC is the sponsor of the I-SPY family of trials: I-SPY2 TRIAL, DCIS RECAST, I-SPY COVID Trial and I-SPY Phase 1b. The I-SPY 2 TRIAL (Investigation of Serial studies to Predict Your Therapeutic Response with Imaging and Molecular analysis) is the longest running platform trial, continuing its operations for over 10 years. QLHC has extensive experience building and managing coordinating centers and executive steering committees for scientific direction and program governance. There are approximately 40 trial sites in the QLHC network including many major academic centers and major healthcare providers.

Based on our experience with OneSource and I-SPY COVID, we provide comments below on the specific "Information Requested" topics, including technical, standards, and operational perspectives. Under 'pilot or demonstration project', we outline a framework of phases for approaching the data collection challenge: pre-emergency (pre-implementation and implementation), inter-emergency, early emergency, trial implementation, and during emergency; and discuss what steps should be taken in each phase.

II. Information Requested

1. United States Core Data for Interoperability (USCDI)

USCDI FHIR-based resources allow uniform data to be extracted across EHRs from participating trial sites. Rollout to sites is expedited since certified EHRs are mandated to have USCDI mappings already in place. In the I-SPY COVID trial, automated acquisition of study protocol-required data elements via available USCDI resources enabled rapid implementation and 60% savings in data collection time and effort. We recommend expansion of USCDI to include additional data elements, including those in the current draft USCDI V4 specification and fields currently in 'Comment' level status such as Adverse Events, Research Data, and Provenance. We also recognize that it is unlikely that all fields required for effective research can be rapidly and uniformly mandated through USCDI, especially early in an emergency when outcome measures are evolving. A hybrid approach relying on USDCI (and other widely adopted FHIR resources) as much as possible, combined with means to easily collect trial-specific data at or near the point of care, with minimal burden on sites, is desirable.

2. HL7 FHIR APIs

Continued expansion - Similar to topic 1, the continued expansion and adoption of structured FHIR resources in areas where unstructured data is now the norm (clinician notes, pathology notes) will benefit research.

Bulk FHIR could allow for easier data collection for chart reviews or population health studies. It can also support site feasibility and eligibility determination/recruitment, (i.e., searching through a population of patients that have a history of heart failure). Bulk FHIR adoption is not currently as broad or mature as traditional FHIR APIs.

FHIR Questionnaire and QuestionnaireResponse offer potential to reduce patient burden, by eliminating duplication for participant reported outcomes (ePROs) - if a patient portal requests a standardized instrument and a trial-specific ePRO system requires the same instrument for the patient on the trial, the EHR-based patient portal should be able to share that data. However the limited semantic capabilities of these FHIR resources makes this difficult to do, and is an area for further development by standards organizations.

Alignment with existing clinical research data standards and models for analytics and regulatory submissions is important. Compatibility of FHIR resources with research standards such as Clinical Data Interchange Standards Consortium (CDISC) and the Observational Medical Outcomes Partnership (OMOP) standards can ensure that EHR data can be easily be incorporated into regulatory

submissions. Existing mapping and crosswalk efforts, such as the FHIR to CDISC Joint Mapping Implementation Guide, should be utilized and extended.

3. SMART on FHIR APIs

SMART on FHIR minimizes implementation burden – it allows third party 'apps' to be launched from an electronic health record with integrated user authentication, context, and API-based data sharing in a highly standardized and portable way. SMART on FHIR is well supported by major EHRs, allowing implementations to be rolled out quickly across sites with minimal variation and low implementation burden. Use of SMART on FHIR should be a centerpiece for automating emergency clinical trial data collection and expanding clinical research into underserved settings. It reduces the burden of research participation for institutions that have limited IT resources, with shorter implementation times and minimized privacy, security and implementation risk to sites.

Integrates into clinical care without disrupting existing workflows – it enables launch points from directly within the patient chart, providing a highly integrated way to collect research data, with minimal disruption to existing clinical workflows, ability to capture standardized elements, and augment those with protocol specific requirements.

Proven in production use - As part of OneSource, OpenClinica and Quantum Leap have deployed the OpenClinica Unite SMART on FHIR app to 15 large hospitals and health care networks for collecting data in multiple multi-site interventional clinical trials. SMART on FHIR has been critical to the success of those integrations. Traditional EHR integration projects are difficult to prioritize due to the resources required. With OneSource, when it was clearly communicated to sites that the integration was SMART on FHIR-based, site IT teams understood that the effort would require minimal resources and be highly systematized, allowing projects to move to the top of the priority queue. In practice, we have found each integration takes less than 15 hours of site technical time to complete.

Allows for re-use and expansion, and hybrid data collection models - Once an app integration is in place, the data collection can be configured for re-use across multiple studies. When a record is completed, data can be sent to the research database and, when needed, written directly back to the EHR. Due to its accessible launch point within the EHR, thes apps can also be used to deliver additional features such as "how to" videos, protocol training resources, and patient educational materials.

4. Clinical Decision Support (CDS) Hooks

Clinical Decision Support (CDS) Hooks offer an exciting opportunity to automate eligibility determination, reporting of adverse events, and assisting with the order of events for a clinical trial. Adoption of CDS Hooks between EHRs and other systems is still in early development, so infrastructure would need to mature before the specifications could be more widely used for a clinical trial. Care should be taken to minimize alert fatigue on users.

5, 8. Operationalizing protocols of varying complexity and Capturing data elements required for clinical trial protocols

Hybrid of automated and facilitated data collection models - FHIR data is crucial in expediting data collection as part of a clinical trial. It reduces the burden, complexity, and delays by automating collection of structured data and minimizing manual data entry. The complexity and specificity of most studies, while relying greatly on FHIR and USCDI. will also require more targeted collection of trial-specific data elements, currently done through electronic Case Report Forms (eCRFs) or Direct Data Capture (DDC) methods. Through SMART on FHIR, sponsor-managed data collection apps can be launched at or near the point of care from the patient chart and can surface these data collection instruments, potentially combining them with data acquired from the EHR, ePRO, and other sources to offer providers real-time operational logic and decision support, and to minimize data collection burden.

Minimize logistical & IT burden on sites - Having the apps and CRFs/data definitions be managed by the trial sponsor (or their designated provider) rather than natively within the EHR drastically streamlines change management should the protocol change, as site IT personnel will not need to be involved in deployment of amendments.

Avoid duplicate documentation in different systems - Potential to write back data from the structured eCRF to the patient record exists in current standards, and should be employed where appropriate to eliminate dual documentation burden (ie when data is captured as source in the research system).

6. Consent, deidentification, return of results.

Electronic Consent - Using electronic tools as part of the consent process can make it an engaging, accessible, and informative experience for patients and caregivers, and enable both in-person and remote consent. It is crucial to ensure all electronic records are 21 CFR part 11 compliant while still being user-friendly and accessible to patients and their caregivers. This compliance requires clinical

trial software to make certified copies, maintain audit trails, and archive records, as well as lays out standards for electronic signatures.

Deidentification - Regarding deidentification and managing protected health information (PHI), FHIR APIs can be designed to exclude PHI by determining which elements are to be integrated. In OpenClinica's experience, EHR FHIR Resource IDs can be used to establish and maintain research participant ID to EHR patient linkage without exposure of PII to researchers/sponsors .

7. User interface and experience.

Simple, straightforward data entry optimizes user experience. Accessing data directly from the EHR to the EDC, launched using Single Sign on from the EHR, is timelier, less error-prone, more secure, and reduces staff burden compared to re-entering data manually from the EMR to the EDC (known as "swivel chair interoperability"). Additionally, because the data is pulled directly from EHR, data does not need to be source data verified. Data captured directly as source from clinicians can then be written back to the EHR, minimizing the need for dual documentation by busy clinicians.

10. Emerging technologies

Integrated real time data with decision support - Modern research data collection systems can be used to provide real-time operational logic, decision support, and patient safety information for on-study participants at the point of care. Doing this effectively requires accessible launch points from the patient chart, and availability of multiple streams of data on the participant (EHR, ePRO, eCRFs, wearable/home health devices, labs, etc.) in near real time. Implemented thoughtfully in traditional or decentralized trial settings, these systems can help advance science and patient care, while minimizing disruption to existing clinical workflows.

Patient-directed data sharing - Enabling patients to support sharing of their medical record data and other health data is an important emerging area that should be included in this initiative.

11. Pilot or demonstration project

For pilot projects and future wide-scale implementation, the government should utilize and invest in existing networks utilizing automated data collection, to build on existing successes and reduce the effort of scaling to future sites. After investing in networks of sites (and in cross-network interoperability), it will be important to support regular use in practice to sites to ensure they maintain their operations so that studies can be activated quickly at scale when needed.

A pilot project should be organized around milestones that would be needed at different stages of an emergency: pre-emergency, inter-emergency, early emergency, trial implementation, and during emergency. Funding to support pilot projects, simulations, standards development, interoperability demonstrations, and perhaps most importantly, real implementations in live, non-emergency clinical trials will help maximize the value delivered in a future emergency.

Pre-emergency can focus on building operational networks with a common, SMART on FHIR based app architecture, and using them in a pilot/simulation project as well as in real (non-emergency) clinical trials. During this period, defining and implementing standards including new USCDI & FHIR data elements, interoperable digital protocol definitions, and best practices for automated and semi-automated data collection workflows, along with training, capacity building, and implementing determined requirements and forms. Interoperability should be established between networks (e.g. site networks using different SMART on FHIR apps, sponsors using different EDC systems and other research data repositories). Use of existing standards such as CDISC, and advancing interoperable digital protocol definitions will support this goal. Metrics should be established/evaluated ensure study startup and conduct can be performed at an accelerated pace, with a high level of data quality. Additionally, master agreements should be established at sites, especially with integrated delivery networks (IDNs).

During the inter-emergency phase, sites would begin capturing relevant USCDI data and registry data for upload to a central repository, while the oversight body would run regular monitoring reports and conduct registry studies as needed.

Early emergency - Data collected during the inter-emergency would be a resource for the early emergency phase, supporting definition of new outcome measures and providing a baseline of data for related conditions, complications, and safety. Site feasibility analyses can be performed and interaction points for recruitment and eligibility determination can be implemented

Trial implementation, similar to the early emergency phase, would be seamless and quick, as training and form requirements were completed in the pre-emergency phase and can be supplemented through electronically delivered protocol-specific training. As protocols are finalized, calendars with automated scheduling of trial activities delivered through the platform ensure the right procedures are being performed and data is being captured in a timely fashion.

During the emergency, clinical staff would recruit patients, ensure protocol adherence with help from the automated schedule of activities, and collect data as per the pre-emergency clinical data workflows, with oversite of performance based on metrics defined in the pre-emergency phase.

12. Specific commercial capabilities

OpenClinica and QLHC have an existing automated data collection network (OneSource) that is expanding its use in platform trials and other studies. This network has been recognized by BioIT World and the FDA. It automates the capture of trial data from Electronic Health Records (EHR) systems for clinical trials. OpenClinica Unite, the technology powering OneSource, is used across multiple studies to capture regulatory-grade data from Epic and Cerner EHRs. The OneSource/ OpenClinica Unite system can interoperate with other EDC systems and research data repositories. It provides research site personnel with user-centric EHR-integrated workflows and automated data acquisition, minimizing many of the inefficiencies and quality risks of today's 'swivel chair interoperability' practices. The eCRF can be launched from the patient chart with a single click. Structured data from the EHR populates the eCRF, with workflows for user review/validation. eCRFs that cannot be directly populated are accessible for manual entry. Mobile support enables direct data capture at the point of care. The EDC's proven 21 CFR part 11 compliant features support regulatory-grade evidence of data integrity.

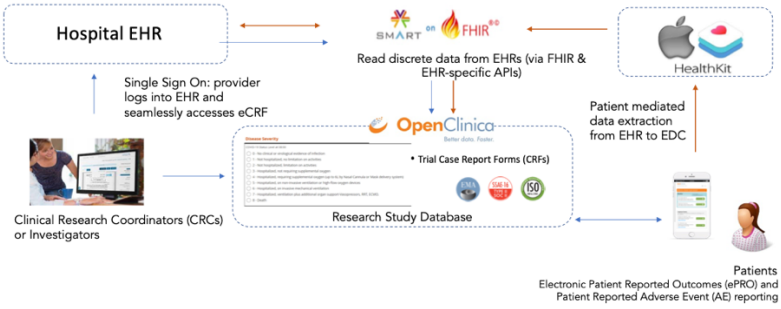


Figure 1: SMART on FHIR integration through 1) on site, institution configuration with the EHR and 2) EHR access points for patient mediated access. Both approaches result in high quality, efficient transfer of discrete data from the EHR to the study database system.

It also enables trial participants to directly share their health data their study they are enrolled in through mobile app-based health record integration (Figure 1). This provides an alternate method of acquiring health record data in which patients authorize sharing of the laboratory and concomitant medication data from their EHR to OpenClinica.

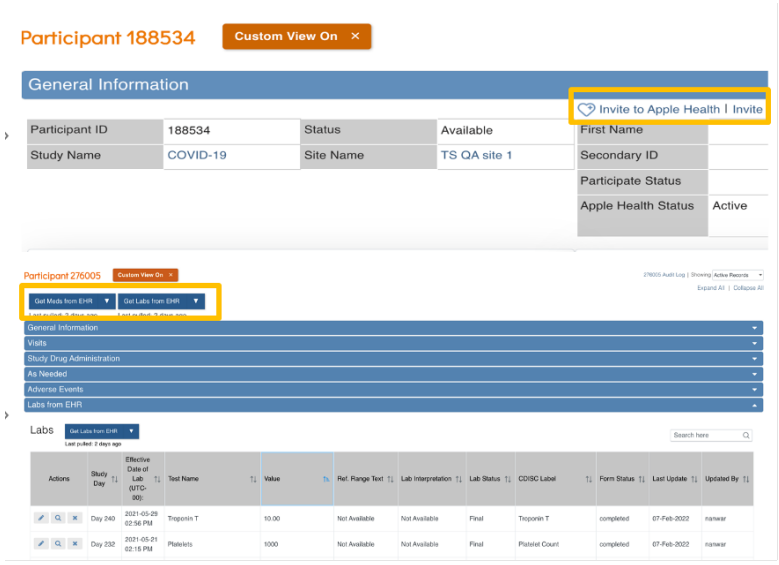


Figure 2: OneSource interface showing the two approaches to pull laboratory and concomitant medication data: 1) Invite patients using patient mediated data access for sites with Apple Health framework for patients to initiate access or 2) CRC's selecting "get Labs" and "get Meds" for sites with OneSource EHR on-site configuration.

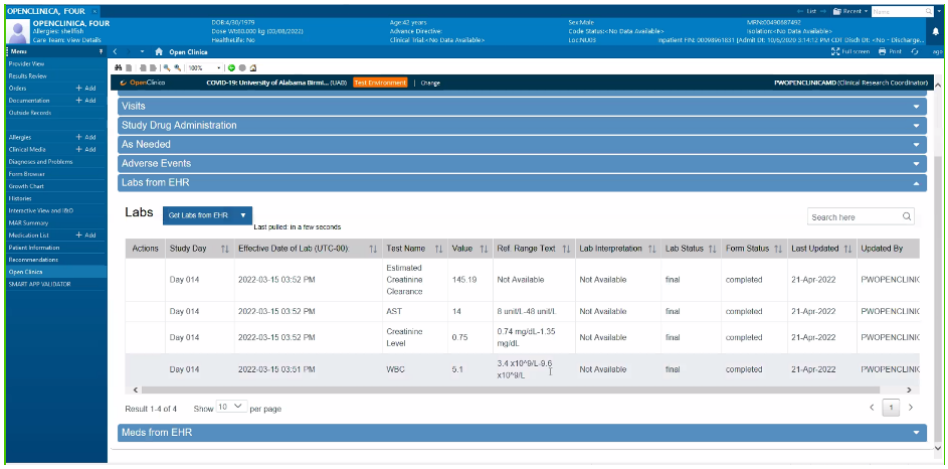


Figure 3: Display of lab data pulled into EHR using patient mediated access configuration or the direct EHR integration set up.

/thoughtworks

Thoughtworks

Request for Information on Data Collection for Emergency Clinical Trials and Interoperability Pilot

Office of Science and Technology Policy (OSTP)

Thoughtworks response to 87 FR 65259
January 9, 2023
Attention: Grail Sipes at 202 -456 -4444; datacollectionforclinicaltrials@ostp.eop.gov

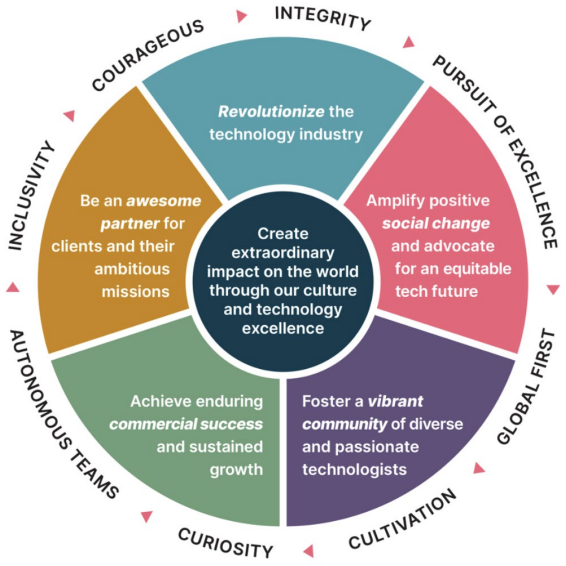
Table of contents

1. Company Information

Name of Company	Thoughtworks Federal
Address	200 E Randolph St., 25th Floor, Chicago, IL 60601 - 6501
Point of Contact	Jude R. Soundar Bryan Nice
Phone Number	773-818-7531
Email address	jude.soundar@thoughtworks.com bryan.nice@thoughtworks.com

2. Relevant Thoughtworks Capabilities

Thoughtworks is a global technology consultancy that integrates strategy, design and engineering to drive digital innovation. We are 12,000+ people strong across 48 offices in 18 countries committed to our core principles and values at right. Over the last 25+ years, we've delivered extraordinary impact together with our clients by helping them solve complex mission and business problems with technology as the differentiator.



Our approach is built upon decades of hands-on experience, paired with deep technical expertise in the areas that are core to every enterprise's technology strategy. We help our clients create and connect strategy to practical execution, using cross-functional teams of industry-leading domain experts, product managers, developers, data engineers, and experience designers.

We enable clients to rapidly and successfully navigate their digital transformation journeys with an unparalleled range of skills and unmatched expertise in digital strategy and operations; data strategy, engineering and analytics; enterprise modernization, platforms and cloud; and customer experience, product and design.

The work we do	
Digital Transformation & Operations *Evolve your organization and deliver digital transformation that increases your agility, resilience, and ability to deliver more mission.*	•Technology strategy and alignment •Agile fundamentals •Value-driven portfolio management •Delivery transformation •Digital fluency
Enterprise Modernization, Digital Platforms, and Cloud *Modernize your operations, platforms, development and deployment practices to consistently deliver greater customer value.*	•Cloud migration •Digital platforms •Delivery Infrastructure •API architecture
Customer Experience, Product, and Design *Drive adoption and improved service delivery through differentiated, customer-centric, digital experiences.*	•User experience design •Product organization transformation •Product delivery and evolution
Data and Artificial Intelligence *Maximize the business value and enable better decisions with pragmatic data strategies, governance, and technology ecosystems.*	•Data strategy and governance •Data platform modernization and engineering •Data Mesh | Data as a Product •Continuous Delivery for Machine Learning (CD4ML)
Agile Culture & Delivery Mindset *Industry leading software development excellence that leverages patterns and best practices for efficient and continuous deployment.*	•Continuous Integration / Continuous Delivery (CI/CD) •DevSecOps •Open source first •Evolutionary and adaptive architectures

Who we work with

VA · U.S. Department of Veterans Affairs · GOV.UK · USAID FROM THE AMERICAN PEOPLE · OKLAHOMA Employment Security Commission · Government Digital Service · NIH · NYU Langone Health · DigitalVA · Australian Bureau of Statistics · VHA Innovation Ecosystem · U.S. DIGITAL SERVICE · MySense · unicef · Partners In Health

Our cloud partners

aws · Google Cloud · Microsoft Azure

3. Response to Questions

1. United States Core Data for Interoperability (USCDI).

USCDI+ should accommodate the list of clinical note data classes enumerated below:

- Bone Marrow (Biopsy/Aspiration)
- Cardiac Catheterization
- DLCO - Diffusing Capacity of Lung for Carbon MonoxideStudy
- Echocardiogram result/interpretation
- Emergency Room Notes
- Genetic Testing
- Mental Status Evaluation
- Neuropsychological Testing
- Operative Notes
- Physical Exam
- Psychological Testing
- Pulmonary Function Study
- Spirometry Test result/interpretation
- Spirometry Tracing Image
- Stress Testing (exercise, pharma)
- Ultrasound (exclude Doppler)
- Doppler Test
- Electroencephalogram (EEG)
- Electromyogram/nerve conduction (EMG)
- Angiogram
- EKG/ECG result/Interpretation
- Myelogram
- EKG/ECG Tracing Image
- Colonoscopy
- Endoscopy
- Audiograms
- Visual Acuity
- Visual Fields
- Holter monitor
- Doctor to Doctor

Clinical notes are critical for determining the efficacy of a clinical trial. Further, these notes are also critical in making accurate disability determinations for the Veterans Administration and the Social Security Administration.

2. HL7 FHIR APIs. a. Use of an API that supports FHIR Bulk Data Access to support clinical research; whether bulk data exports from EHR systems can be used to support certain clinical trial protocols.

> This will be useful to support submissions required to support the clinical study. FHIR Bulk Data Access enables principal investigators to fetch data sets related to their study recruitment efforts. It can also support bulk data submissions back to CMS or FDA in a timely manner. Providing an interface to securely support interoperability will provide greater transparency on clinical trials efficacy.

b. Use of the FHIR Questionnaire and QuestionnaireResponse resources to support clinical research.

This can be used to support surveying recruited participants, site coordinators, and principal investigators. The data can be associated with the study being conducted and reduces latency to have timely data from a clinical trial.

3. SMART on FHIR APIs:

a. The most promising ways to create SMART on FHIR technologies that are portable across different institutions and EHR systems, but also provide adequate functionality to support emergency clinical trial research.

Mobile apps should be the focus to enable SMART on FHIR technologies for emergency clinical trial research.

b. Whether the portability of SMART on FHIR tools provides a way to reach institutions and sites that have limited information technology resources; any promising ways to use SMART on FHIR to expand clinical research into underserved settings.

Web portals and mobile apps used to facilitate submission from areas with underserved settings. The web portal can implement and enforce the FHIR message specifications, without the local IT needing to build a custom application to facilitate data submissions/fetching.

4. Clinical Decision Support (CDS) Hooks:

This can be used to support approval workflows requiring attestation or managing tasks associated with the status of a study. Use CDS Hooks to s upport actions driven by the appropriate human study context at the correct time within a clinical study flow. It will help support facilitating studies to achieve efficacy targets.

5. Operationalizing protocols of varying complexity.

6. Consent, deident ification, return of results.

a. In light of this, we seek comment on how the tools described above can be used to obtain, collect and/or manage any required informed consents and/or authorizations from patients or individuals in accordance with applicabl e regulations.

Associate the research proposals with IRB approval to the clinical study being conducted. Each human subject recruited associates a signature GUID, using smart contract technology, representing the acknowledgement of being informed, giving consent, and authorizing the principal investigator to collect and use their data.

b. We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information.

For this, it is recommended to use OM OP as the data repository standard

https://www.ohdsi.org/data -standardization/. OMOP has an approach to preserving privacy in its implementation https://ohdsi.github.io/CommonDataModel/cdmPrivacy.html.

c. Ideally, patient authorization would allow clinical trial data to be used for additional research beyond the original study.

Use the OMOP standard to create and manage the data registry. https://ohdsi.github.io/CommonDataModel/index.html

d. We seek comment on any technical capabilities that could support return of results to study sites or participants, where appropriate.

Architecturally, there needs to be people who have the technical skills that understand how to design and implement distributed and data intensive systems. They will need to understand FHIR and CDS hooks to create reference implementation on the data contract design, behavior, and output expectations. It is recommended they understand how to design using microservice methodologies to create systems that are dynamic and scalable as usage demand increases. It is recommended they understand DevSecOps engineering practices and how to roll deployments using methodologies like blue/green deployment to support high resiliency.

e. We seek comment on any regulatory or ethical guidelines that are relevant to patients' consents and authorizations under the use case described in this RFI, and on ways in which technical solutions might help ensure adherence to applicable regulatory or ethical guidelines.

Integrate in the workflow for an IRB assessing research proposals and approving can be instrumented to support transparency and adherence. The other aspects can be around the handling of data submissions and approval by the principal investigator or their delegate.

7. User interface and experience. With all of the above technologies, we seek input on:

a. The best way to optimize the experience of health care providers, administrato rs, and other users, so as to maximize the utility and uptake of the product.

API interfaces to support seamless integration within the researcher's organization if they have an IT department. If their research organization or research does not have an IT department, the web portal to support the data movement activities.

b. To the extent a particular form, app or other tool requires input from a health care provider or other user, the best ways to increase the likelihood that users will actually provide th at input. It would be helpful to receive comments on methods that are available for completing empty fields after the fact, or otherwise managing any missing data.

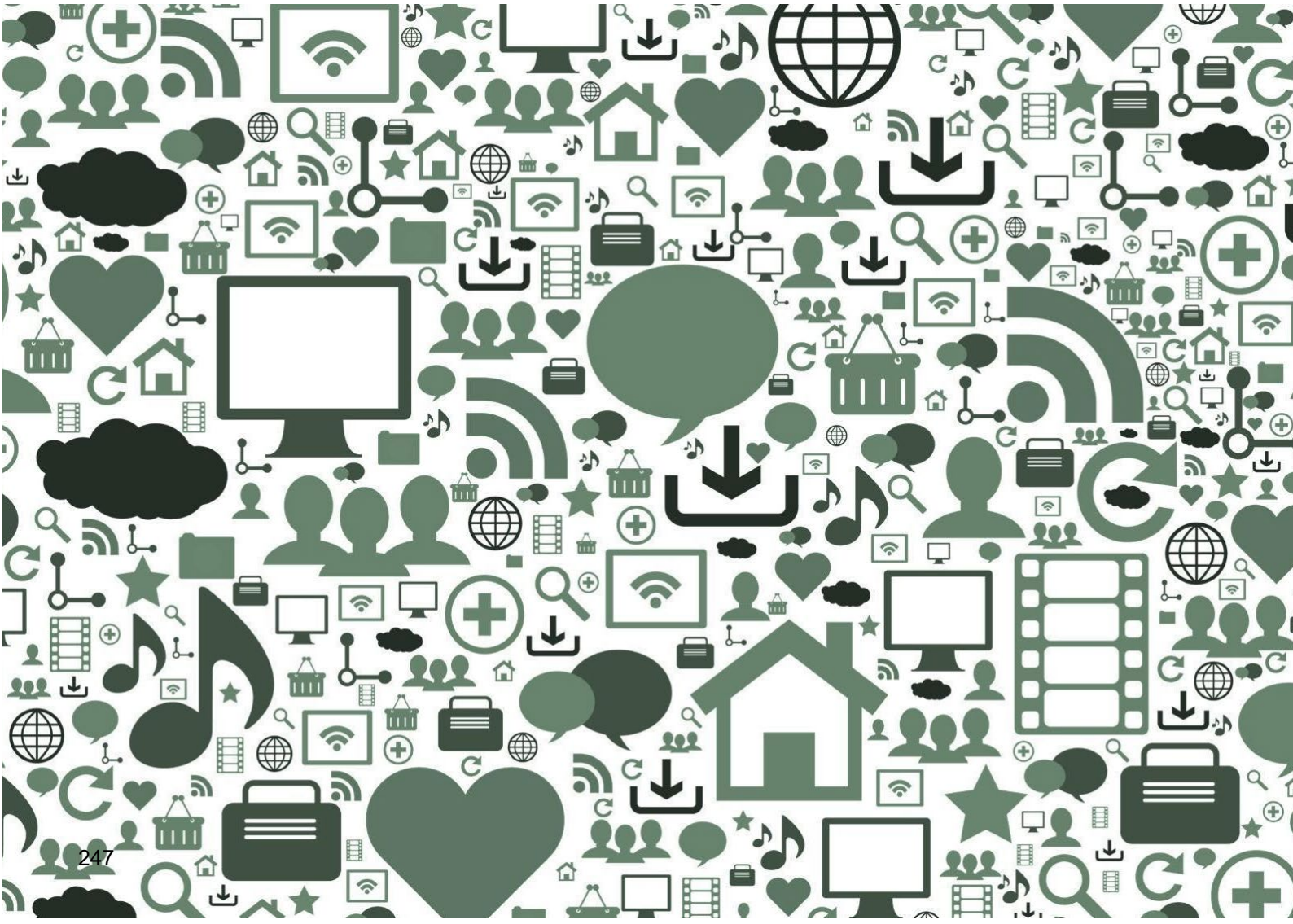
Form verification on required fields. For bulk data submissions the intake processing

method can assess the profile of the data and based on missingness thresholds trigger notification if it was rejected or not. The key will be identifying the target threshold that minimizes bias of data due to missingness.

c. For clinicians and health IT users: what existing tools, apps, or processes you have found most usable and why.

From a data science perspective, python and R connecting to any API tends to be the preferred approach.

/thoughtworks



Submitted by Laura Marcial, PhD, FAMIA, RTI International

RTI International is an independent, nonprofit research institute dedicated to improving the human condition. Our vision is to address the world's most critical problems with science-based solutions in pursuit of a better future. Clients rely on us to answer questions that demand an objective and multidisciplinary approach—one that integrates expertise across the social and laboratory sciences, engineering, and international development. RTI experts in informatics and clinical trial coordinating center management put together the responses to this RFI.

1. United States Core Data for Interoperability (USCDI).

At a foundational level, research and clinical trials need to have increased awareness of USCDI data classes and assess early in the design phase how electronic health records (EHRs) and other types of health IT–generated data would likely be used in studies or trials. Speakers at the January 11, 2023, Office of Science & Technology Policy/Office of the National Coordinator for Health Information Technology (ONC) listening session shared that researchers and study designers must change their approach and build with existing EHR data captured, their related terminology standards, and methods for exchanging those data (e.g., Fast Healthcare Interoperability Resources [FHIR]). This would provide value in multiple ways: (1) prevent creating one-off data collection systems that cannot be integrated into the workflow and systems used by clinicians; (2) allow use of technology to support sharing of information between a clinical site, a patient, and a research entity; and (3) reduce administrative burden. Using the framework of USCDI and USCDI+ provides researchers the ability to identify where there are gaps that need to be addressed and whether those gaps can be addressed by proposing new data elements in an existing class, proposing new use cases and value sets for existing FHIR implementation guides, or highlighting the need for a new extension for USCDI clinical trial data.

Existing work could help inform future extensions of USCDI to support emergency clinical trial research. The Clinical Data Interchange Standards Consortium (CDISC) has defined a number of standards that support the capture and sharing of information related to research and clinical trials. In a joint effort with Health Level Seven (HL7), CDISC created a FHIR implementation guide that defines mappings between FHIR and three specific CDISC standards: the Study Data Tabulation Model, Clinical Data Acquisition Standards Harmonization, and LAB. Through this work, a few challenges have been identified. For example, FHIR does not provide a way to identify study versus nonstudy data elements such as laboratory tests, concomitant drug use, and so on, and does not have a specific mechanism to capture events, such as AdverseEvents, that could occur but have not yet occurred. QuestionnaireResponse could be used to capture these data elements but would benefit from a more standardized approach.[1]

Extensions for specific use cases, such as oncology, may also be needed. As part of the CodeX HL7 FHIR Accelerator, the ICAREdata project leverages minimal Common Oncology Data Elements to collect key outcome data that were not typically captured from the EHR in a structured way, such as cancer disease status and treatment plan changes. The project also serves as a pilot for sharing clinical trial outcome data from select National Clinical Trials Network institutions for oncology research.[2]

2. HL7 FHIR APIs.

a. Use of an API that supports FHIR Bulk Data Access to support clinical research; whether bulk data exports from EHR systems can be used to support certain clinical trial protocols
In spring 2020, ONC published a rule regulating the 21st Century Cures Act application programming interface (API) requirement along with protections against information blocking.[3] One of the APIs covered in this rule is the Substitutable Medical Applications, Reusable Technology (SMART)/HL7 FHIR Bulk Data Access API, or Flat FHIR API, which enables access to patient-level data across a population, supporting many use cases across the healthcare ecosystem, including clinical trials.

Limitations to implementing bulk FHIR noted in a recent study[4] include implementation hurdles such as (1) hardware limitations and logistics for moving large datasets, including error handling, processing time, and deciding where to split large files and how best to load them; (2) managing granular access, particularly in federated systems and where the user requesting a bulk export needs to be explicitly identified for audit purposes; and (3) de-identifying data stored in documents and free-text fields when leveraging the exported data for some use cases.

Although this approach holds a lot of promise, there are still technical details to be worked out.

b. Use of the FHIR Questionnaire and QuestionnaireResponse resources to support clinical research.

On principle, the use of questionnaire and QuestionnaireResponse are flexible, highly adaptable, and extensible. There are great tools to leverage, such as LHC-Forms, to develop and implement assessments to capture patient-reported outcomes and other trial data. However, this capability lacks sophistication for clinical trial use, often requires adaptations to support the analysis side of trial work, and is only in somewhat limited use today.

3. SMART on FHIR APIs:

a. The most promising ways to create SMART on FHIR technologies that are portable across different institutions and EHR systems, but also provide adequate functionality to support emergency clinical trial research.

The ONC Cures Act Final Rule mandating access to FHIR APIs has created a ripe environment for app services that support patient or consumer access to data. There is an increasing precedent for access to cloud-based services that could support a “write once, deploy everywhere” solution for data collection using bulk FHIR. There are also efforts such as CDS Connect, a repository of interoperable shareable clinical decision support (CDS) applications or solutions, many of which are FHIR based. Efforts such as these may go a long way toward fostering the kind of services needed for emergency clinical trial use.

b. Whether the portability of SMART on FHIR tools provides a way to reach institutions and sites that have limited information technology resources; any promising ways to use SMART on FHIR to expand clinical research into underserved settings.

The widespread use of EHR vendor systems, at least for hospitals and increasingly for smaller clinics in all settings, is a disruptive source of access to patient populations including those in underserved settings. Additionally, where those settings are utilizing compliant vendor systems with FHIR APIs available, use of FHIR applications to support interoperable data transmission is viable. In short, if a small rural hospital is running an Epic or Cerner EHR system, that hospital should have access to a FHIR API and that FHIR API should have a publicly available endpoint from which any compliant, registered FHIR application can retrieve data with the appropriate access.

4. Clinical Decision Support (CDS) Hooks:

Although ONC's Cures Act Final Rule has improved the promise of interoperability, and much progress has been made to support data exchange and to demonstrate that support, gaps still exist between what USCDI mandates and what is needed to develop useful, functional, and usable SMART on FHIR applications.

In late 2019, a team led by RTI began an effort to develop and implement a publicly shareable provider-facing SMART on FHIR CDS application called PainManager to facilitate the management of chronic pain. Included in this application is a calculator, based on the Centers for Disease Control and Prevention Opioid Prescribing Guideline, that uses a patient's current opioid medications to calculate morphine milligram equivalents per day for the provider to review in the context of a shared decision-making encounter. This information is then used to facilitate the development of a chronic pain management plan with the patient.

Implementing the application required functionality beyond the native FHIR API provided by the EHR to integrate the calculator into the provider-facing app and persist patient-reported data. The project also exposed policy-related challenges in accessing prescription drug monitoring program data and led to important lessons learned related to integrating alerts into clinicians' workflow.

The above example highlights how CDS can be integrated into the vendor system environment and incorporated into the clinical workflow. CDS Hooks is a service that most compliant EHR systems have integrated and offers a lot of promise for triggering events within the system and the clinical workflow, but to date, effective use of CDS Hooks has been very limited. Its limitations are likely due in part to organizations not having a great deal of experience leveraging CDS Hooks in their vendor environment.

5. Operationalizing protocols of varying complexity.

a. Whether any of the tools described above might be particularly well suited for certain types of studies.

b. For example,

i. *Whether a bulk FHIR API export could be used to gather data for a simple trial protocol that is relatively close to the standard of care for a particular condition.*

Given greater compliance with USCDI and increasing capacity, especially at academic medical centers, a bulk FHIR export could support data gathered for a simple trial protocol in the near future. Today, the limitations appear to be more limited to policy and governance mechanisms that make this work a challenge. Before considering bulk FHIR API export, the underlying data classes and data elements must be in place. For example, even with very simple, pragmatic trials, required data extraction can be complex.

ii. Whether a FHIR Questionnaire/QuestionnaireResponse or a SMART on FHIR form would be useful in capturing data for a more complex protocol, such as one that involves an investigational agent.

Complications during even a simple trial can make automated data extraction unmanageable. Because consistency is critical, the same value needs to be measured in the same way across all sites, so establishing consistency and standardizing data capture (e.g., for blood pressure readings) for the clinical environments must happen first. It is important to stress that clinical environments are complex and inconsistent in terms of standard of care and so on. Data capture is just one component of a clinical trial. One approach that might be a good starting point is to look at something like automated capture of safety data. The data are more routine, more generally consistently captured, and stored in the EHR, which makes them good candidates for a preliminary pilot. Eventually, these data could be used to augment existing processes (e.g., electronic data capture [EDC] system-based data capture) until more is known or understood about needed improvements in data capture. Creating a roadmap for this work would be ideal.

c. Any technical limitations that we should be aware of regarding use of the above tools to operationalize clinical trial protocols.

The FHIR bulk $import operation is still under development and not currently a consistent approach to ingesting the data available from the bulk $export operation. Loading of FHIR resources requires checking for referential integrity and merging and synchronizing the data. These operations could lead to performance issues when done at scale and need to be further explored.[5] Moreover, the community is still developing the tooling required for common tasks such as de-identifying patient data, mapping between terminologies, and filtering the data.[6]

Challenges related to data governance also must be overcome. Many EHR vendors limit the ability to persist data from outside the EHR. Part of this limitation is due to privacy and security concerns, so any solutions would require enhanced security. RTI is leading a pilot of a SMART-on-FHIR app that seeks to overcome these hurdles by using a research data store that sits between the app and the EHR to store the data gathered from the app (patient-reported outcomes) and multiple EHR sources. These types of proof-of-concept pilots are really needed to fully determine the challenges and potential solutions for many of the concepts laid out in this RFI response.

6. Consent, deidentification, return of results.

a. In light of this, we seek comment on how the tools described above can be used to obtain, collect and/or manage any required informed consents and/or authorizations from patients or individuals in accordance with applicable regulations.

The FHIR Consent resource can be utilized to identify the date and source of the patient's consent, the time frame for consent, the scope (e.g., research, treatment), actions authorized by the consent (e.g., collect, retrieve, and disclose patient information), the level of security (e.g., unrestricted, substance abuse, HIV/AIDS, sexual and reproductive health), and the purpose of the consent (e.g., healthcare delivery, coordination of care, clinical trials, public health research). This resource also allows for adding additional consent or removing consent.[7]

The FHIR Consent resource has been successfully used in the *Nutrition for Precision Health*/*All of Us* Research Program by researchers at RTI. Using FHIR R4, there were limitations in the ability to capture the site where the consent occurred via the Consent resource, but these limitations have been addressed in FHIR v5.[8]

b. We also seek input on what additional capabilities would be required to deidentify or otherwise manage protected health information. It would be helpful to receive comments on which deidentification and protection approaches are sufficiently mature to support a pilot effort in the near term.

Microsoft has developed some open-source tools for health data anonymization intended for use cases in research and public health. The tooling is built into the bulk $export operation for Microsoft's open-source FHIR server for Azure and can de-identify or redact patient information, based on a configuration file, before export. Most FHIR servers do not have this capability. The tooling could also be used in a data pipeline but would require transfer of identifiable patient information, then de-identification before ingestion into the registry.[9]

c. Ideally, patient authorization would allow clinical trial data to be used for additional research beyond the original study. We would appreciate input on how the content collected for consent and authorization as well as the interfaces with deidentification technologies should be designed to enable flexible and responsible reuse of clinical trial data.

Ideally, the FHIR Consent resource could be leveraged to limit the data returned. The data elements in this resource include the scope (e.g., research, treatment), actions authorized by the consent (e.g., collect, retrieve, and disclose patient information), the level of security (e.g., unrestricted, substance abuse, HIV/AIDS, sexual and reproductive health), and the purpose of the consent (e.g., healthcare delivery, coordination of care, clinical trials, public health research). The ResearchStudy resource could be used in conjunction with the Consent resource to return the appropriate data when bulk data are requested.

The FHIR bulk data API search parameters can be used to identify resources based on specific criteria (e.g., diagnosis, medication utilization) but could use further refinement and the ability to limit the resources returned beyond by type (e.g., Patient, Encounter, Observation, Dispense) and elements for a resource type, which is experimental in the Bulk Data Access Implementation Guide 2.0.[10]

d. We seek comment on any technical capabilities that could support return of results to study sites or participants, where appropriate.

FHIR-based applications can be leveraged to provide at least a review of information submitted (assessments, etc.) by the study participant. When integrated into patient portals, these apps can be used to complement other portal functionality to provide updates and messaging around trial recruitment and results. FHIR services combined with Clinical Quality Language (CQL) can provide calculated or logic-based solutions (e.g., electronic quality measures and other reporting). Where FHIR servers are communicating bidirectionally, this information could be supplied at the site level or to the patient or participant.

e. We seek comment on any regulatory or ethical guidelines that are relevant to patients' consents and authorizations under the use case described in this RFI, and on ways in which technical solutions might help ensure adherence to applicable regulatory or ethical guidelines.

Consumer-mediated data exchange may offer researchers a way to acquire the EHR data they need without confronting these logistical barriers. There are two approaches. In one, which we call Download and Send, study participants use a consumer-facing app to download and aggregate their own health records, which they then contribute to the research database. In the other, which we call Transmit, study participants use an app that directs their providers to transmit their data to the research database. The 21st Century Cures Act states that consumers must be able to access their own electronic health information "with no special effort." The MyHealthEData initiative is predicated on the belief that all individuals should have access to their electronic health information and be empowered to use it however they wish. To implement these principles, the Centers for Medicare & Medicaid Services (CMS) had incentivized providers to give patients the ability to view, download, and transmit their own data electronically, originally through the Meaningful Use program, and now through the Hospital Inpatient Prospective Payment System. CMS and ONC took another major step with rulemaking intended to accelerate the interoperability of electronic health information in the United States by leveraging consumer-mediated data exchange. ONC rulemaking has made it much easier for consumers to access and use their own EHR data with the assistance of any consumer-facing apps that leverage FHIR APIs.

7. User interface and experience. With all of the above technologies, we seek input on:

a. The best way to optimize the experience of health care providers, administrators, and other users, so as to maximize the utility and uptake of the product.

In general, taking a user-centered design approach to the development and implementation (all aspects of the life cycle of a project) is crucial to success here. Additionally, paying close attention to the use of time and workflow management (reducing redundancy and any extraneous burden) is critical.

b. To the extent a particular form, app or other tool requires input from a health care provider or other user, the best ways to increase the likelihood that users will actually provide that input.

It would be helpful to receive comments on methods that are available for completing empty fields after the fact, or otherwise managing any missing data.

Clinicians and researchers who have participated in prior evaluations of health IT used for clinical care or research have indicated that tools that do not duplicate efforts in their workflow are more likely to be used. Barriers for use include redundant data entry and information review.

c. For clinicians and health IT users: what existing tools, apps, or processes you have found most usable and why.

Leveraging tools that exist in the current workflow would increase use. Having a mechanism to flag required actions within the browser in which care is documented would reduce barriers to use.

8. *Capturing data elements required for clinical trial protocols.*

a. We seek comment on the most promising technical approaches that would leverage common APIs to translate a particular clinical trial's data elements into data elements captured by user-facing tools (e.g., FHIR Questionnaire feeding into a SMART on FHIR form or application).

Technical approaches that may be leveraged include Clinical Data Interoperability Services (CDIS) within the REDCap electronic research data capture platform. CDIS is an advanced feature of REDCap that allows an individual REDCap project to interact with an EHR (e.g., Epic, Cerner) using HL7 FHIR API to pull selected information from the EHR into the REDCap project. Users can map EHR data to REDCap fields and initiate data transfer on their own, and the module supports single or batch data pulls from the EHR. Data elements include Basic Demographics, Labs, Problem List, Medications, Allergies, and Vital Signs.

b. If a tool such as a FHIR Questionnaire, FHIR QuestionnaireResponse, or SMART form or app is used to capture required data elements in this way, we seek comment on whether that creates an effective method for "pushing out" a research protocol to investigators and sites.

Conceptually, this type of data element capture may be possible, but there are limited examples of use of FHIR resources for this purpose. Also, this work is operationally difficult even with the simplest trials.

c. It would be helpful to receive comments on how best to ensure compliance with regulatory requirements for eCRFs when designing interfaces for data capture.

Today, EDC systems handle complex regulatory requirements with the incorporation of audit trails, details around capture and maintenance of an original value, permissions for different actors in the system, and other system access control information (e.g., information that might be contained in the provenance resource in FHIR). Other secure access details and management of blinded access would be required. Attention to 21 CFR criteria including management of electronic signatures and ensuring a validated system and a system that supports management of privacy concerns would be needed. Additionally, existing EDC systems provide some level of

254

data quality services, the ability to lock data, and management of *a priori* controls for capture and compliance with the Federal Information Security Modernization Act. Proposed solutions would need to adequately manage these important regulatory requirements.

9. TEFCA and QHINs.

a. Certain policy and/or technical constraints will need to be specified for currently authorized Exchange Purposes under the Common Agreement (e.g., Public Health). We seek comment on which of these constraints will also be applicable to a future research-focused Exchange Purpose.

The Trusted Exchange Framework and Common Agreement (TEFCA) provides the framework standards that health information networks (HINs) and health information exchanges (HIEs) must follow to share data, as well as the legal agreement that governs data sharing between networks. The short-term implications of TEFCA are that the needed infrastructure TEFCA creates will provide the data-sharing capabilities that the research and clinical trials need to create a diverse biomedical data resource. Participation from provider organizations to become a part of the Qualified Health Information Networks (QHINs) is a key factor. ONC recently announced its first QHIN and is processing other applications. For example, eHealth Exchange, CRISP Shared Services, and NextGen Healthcare have announced their intention to apply. The level of participation not only depends on how many organizations sign up to participate, but also the mandates and other incentives that may be established for organizations to participate as QHINs.

Importantly, TEFCA is centered on patient access, not specifically research. Patients may consent to pass through their information, but relevant HINs and HIEs will need to have a common interface and privileges to pass the information to research entities with agreements in place. In theory, these data coming through such interfaces should be normalized.

The intermediate or long-term implications of TEFCA should be the ability for HINs and HIEs to seamlessly connect with each other. There are still significant limitations for cross-network exchange because the obstacles to connecting such diverse systems are only described theoretically. How quickly research may begin to leverage the data depends on the following:

1. Motivation of health organizations to participate in QHINs: This appears to be significant.

2. Ability of health organizations to rally around a common standard such as FHIR or the Observational Medical Outcomes Partnership (OMOP): This has seen slow, but there is steady progression; this likely will not be in place when the exchange first begins, making effective exchange of discrete data a longer-term goal.

3. Willingness of QHIN participants to make their data available for research purposes: This will likely be regionally based, following individual contracts between QHINs and their constituents. However, with the capability to exchange data nationally established, the barrier has been lowered with respect to aggregating national datasets.

4. Willingness of patients to make their data available for research purposes: In theory, if individual patients are willing, then data could be accessed in a patient-mediated manner, thereby drawing information from the HIEs and HINs indirectly. This would bypass item 3 but would still be subject to other limitations.

c. How the Public Health Exchange Purpose could be used to advance the goals of this RFI; what aspects of the use case described above might fall within the scope of the Public Health Exchange Purpose.

The "Public Health Exchange Purpose" outlined in TEFCA could be leveraged for research under certain emergency use circumstances. However, to be effective at all for emergency use, it must be in place for regular use. What is needed more specifically, though, to establish regular use, is an exchange purpose for "Research."

*d. How a **future research-focused Exchange Purpose** could be structured to advance the goals of this RFI.*

Ultimately, the development of a statement for a "Research" exchange purpose should be pursued with relevant clinical trial contributors. As a minimum, this should include the U.S. Food and Drug Administration, the National Institutes of Health (NIH), members of the clinical research community, members of the EHR vendor community, and patients.

e. Other opportunities or constraints related to TEFCA that should be considered with regard to this RFI.

Permissions management may fall to the patient or consumer and could present an undue burden that further excludes marginalized populations.

10. Emerging technologies.

a. How future technologies might affect the use case and underlying assumptions laid out in this RFI.

We presume that cloud-based, app-oriented FHIR services or support will continue to expand in all relevant domains and that logic-based integration (e.g., via CQL) might also be vendor supported in the future. This expansion would greatly extend the potential of these interconnected, interoperable federated environments for research.

b. How future technologies might change the nature of the software architecture, data architecture, or potential data collection solutions for clinical trials.

Today, the existing SMART-on-FHIR solutions depend heavily on vendor marketplaces and a registration process. Provisioning for data privacy and security inhibits further leveraging of cloud-based services and data sharing at scale in a more open application marketplace. Presumably, TEFCA, expansion of USCDI, and the broader propagation of FHIR services will allow for a more robust app marketplace, extended use of cloud-based services, and more rapid

and more accurate data collection for clinical trials. Taken together, these changes would fundamentally impact the architecture of today's solutions. Some of this is conceptually articulated in the MedMorph Reference Architecture Implementation Guide.[11]

11. Pilot or demonstration project.

a. Whether data can be managed through a central repository or small set of central data repositories; options for cloud-based data storage.

The *All of Us* Research Program uses a central data repository to collect data from a wide variety of sources, including surveys, EHRs, biosamples, physical measurements, and wearables like the Fitbit. The OMOP Common Data Model (CDM) is used to standardize these data for researchers. After harmonizing the EHR data to meet the specifications of the OMOP CDM, the data are processed to ensure participant privacy is protected. Steps are then taken to clean and deliver high-quality data.[12] *All of Us* is exploring use of FHIR to exchange data with and for participants.

The NIH Cloud Platform Interoperability Effort (NCPI) is working toward a federated data ecosystem across NIH platforms. The participating platforms include genomic, biomedical, cancer, pediatric, and biotechnology data. The goal is to make the most effective use of the data managed by NCPI platforms. By allowing users to view, browse, and search datasets available across all resources, this data ecosystem can be used by biomedical researchers to better understand what data are already available. This, in turn, will allow for better experimental design of future studies and will prevent duplication of current and past efforts.[13] The initiative's FHIR Working Group is exploring the potential of HL7 FHIR to support the exchange of clinical and phenotype data among the NCPI effort's participating platforms.[14]

b. Technical options that might hold promise in the short term to enable researchers from diverse locations to analyze the data collected from multiple clinical trial sites. We also seek comment on any additional options that should be considered in the long term.

Tools in the marketplace provide important solutions to different components of the system that are needed to support emergency clinical trial research. One component is the near real-time transformation of EHR data into OMOP using artificial intelligence and machine learning tools. One example of such work is evident in the Piano platform created by Australian technology company Evidentli. Some organizations such as MaxMD are working to develop tools and solutions for higher-quality data capture in the clinical environment. Extending FHIR capabilities and services is another area where development could significantly enhance the ability of organizations to more effectively participate in emergency clinical trial research. Canadian technology company Smile Digital Health is doing great work making HAPI FHIR server infrastructure more usable and accessible to health systems with less in-house capacity to develop in HAPI FHIR natively. Still other organizations are tackling the conversion of data in Consolidated Clinical Document Architecture format to FHIR.

[1] Health Level Seven (HL7). (2021, August 31). *FHIR to CDISC joint mapping implementation guide*. http://hl7.org/fhir/uv/cdisc-mapping/STU1/index.html

[2] Confluence. (2022, July 14). *EHR endpoints for cancer clinical trials*. https://confluence.hl7.org/display/COD/EHR+Endpoints+for+Cancer+Clinical+Trials

[3] SMART. (2019). *SMART/HL7 Bulk Data Access (Flat FHIR)*. https://smarthealthit.org/smart-hl7-bulk-data-access-flat-fhir/

[4] Jones, J., Gottlieb, D., Mandel, J. C., Ignatov, V., Ellis, A., Kubick, W., & Mandl, K. D. (2021). A landscape survey of planned SMART/HL7 bulk FHIR data access API implementations and tools. *Journal of the American Medical Informatics Association*, *28*(6), 1284–1287. https://doi.org/10.1093/jamia/ocab028

[5] Neilley, V. (2021, November 3). Most FHIR servers are unusable in production. *Medium*. https://vneilley.medium.com/most-fhir-servers-are-unusable-in-production-8833cb1480b1

[6] Mandl, K. D., Gottlieb, D., Mandel, J. C., Ignatov, V., Sayeed, R., Grieve, G., Jones, J., Ellis, A., & Culbertson, A. (2020). Push button population health: The SMART/HL7 FHIR bulk data access application programming interface. *npj Digital Medicine*, *3*, 151. https://doi.org/10.1038/s41746-020-00358-4

[7] Health Level Seven (HL7). (2019, November 1). *Resource consent - content*. http://hl7.org/fhir/R4/consent.html

[8] Gulden, C., Blasini, R., Nassirian, A., Stein, A., Altun, F. B., Kirchner, M., Prokosch, H., & Boeker, M. (2021). Prototypical clinical trial registry based on Fast Healthcare Interoperability Resources (FHIR): Design and implementation study. *JMIR Medical Informatics*, *9*(1), e20470. https://doi.org/10.2196/20470

[9] Github. (2022, October 20). *FHIR data anonymization*. https://github.com/microsoft/Tools-for-Health-Data-Anonymization/blob/master/docs/FHIR-anonymization.md

[10] Health Level Seven (HL7). (2021, November 9). *Bulk data access IG*. https://build.fhir.org/ig/HL7/bulk-data/export.html

[11] Health Level Seven (HL7). (2023, January 5). *Making Electronic Data More available for Research and Public Health (MedMorph)*. https://build.fhir.org/ig/HL7/fhir-medmorph/index.html

[12] All of Us Research Hub. (2023). *Data methods*. https://www.researchallofus.org/data-tools/methods/

[13] National Human Genome Research Institute (NHGRI). (n.d.). *NIH Cloud Platform Interoperability Effort*. https://anvilproject.org/ncpi

[14] National Human Genome Research Institute (NHGRI). (n.d.). *Working groups*. https://anvilproject.org/ncpi/working-groups

Comments of
The Health Record Banking Alliance
In response to
Office of Science and Technology Policy (OSTP)
Request for Information (RFI) on Data Collection for
Emergency Clinical Trials and Interoperability Pilot
87 FR 65259 (Oct. 28, 2022)
Submitted on January 25, 2023 via *datacollectionforclinicaltrials@ostp.eop.gov*

The Health Record Banking Alliance (HRBA)[1] offers comments in response to OSTP's Request for Information on optimizing data capture for emergency clinical trials. *Please note: these comments complement, and should be read in conjunction with, HRBA's comments, also filed this date, in response to OSTP's Request for Information on clinical research infrastructure for purposes of conducting emergency clinical trials. Please refer to the Appendix in those companion comments for a schematic of Health Data Banks.*

Overview of HDBs and Clinical Research

As noted in HRBA's concurrently filed comments on clinical research infrastructure, Health Data Banks (HDBs) are secure, private- or public-sector institutions. HDBs will offer secure, encrypted repository accounts that patients and other consumers own and control, and where they can aggregate, store, and analyze their health data. Health data includes (and is not limited to) encounter reports – institutional medical records – at clinician offices and hospitals, pharmaceutical data, and payment information related to health care. This information can be integrated using software at the HDB to create longitudinal, problem-oriented Personal Health Records (PHRs), access to which consumers control.

Personal Health Records (PHRs) have been conceived primarily for applications in health care delivery. However, high-quality, longitudinal, patient-centered data sets in Health Data Banks (HDBs), available with patient consent via new information flows that HDBs make possible le, will be transformative for research as well. As they mature and are networked, HDBs will give researchers actionable, consented access to research-grade data sets, currently beyond practical reach. These data sets will come from aggregated medical record and patient-supplied data ("collectively, real-world data" or "RWD") aggregated and compiled in patients' HDB accounts. New, two-way communication channels enabled by HDBs will facilitate participation by a broad cross section of the public in medical research, including in ambitious projects exemplified by the Cancer Moonshot and the All-of-Us Research Program.

Given the potential comprehensiveness of the patient-specific PHR data sets that HDB PHRs will contain, and the fact that patient accounts can, depending on HDB business models and operating protocols, be permissioned to various extents for research, HDBs can be a data treasure for observational studies, for planning, accrual, and follow-up of interventional studies,

[1] The Health Record Banking Alliance, P.O. Box 6580, Falls Church, Virginia 22040, is recognized as a business league by the Internal Revenue Service under Section 501(c)(6) of the Internal Revenue Code.

and in some cases for use *during* the conduct of randomized controlled trials (RCTs). HDBs can facilitate natural history studies and epidemiological research. They can be a reliable source of data for external control arms. Capacity for executing post-market commitments and post-market evidence-generation can be built in. HDBs may also provide efficient channels for obtaining patient input to trial design and selection of clinical endpoints. The data normalization of PHR data that is possible with HDBs will support use of digital tools in clinical research, as well as advanced computational techniques such as synthetic data, digital twins, and generative AI.

HDBs as Emerging Technology for Clinical Trials Infrastructure

Because HDBs are a standing resource for health care, they are also a standing resource for research. What distinguishes HDBs from other sources of real-world data used today is that many HDB business models will readily support iterative, interactive, two-way flows of information between patients at one end, and researchers at the other. Because this adds confirmatory value, it encourages high levels of trust in baseline patient data sets. It also offers research flexibility in the face of the unexpected. That is, by definition, the very nature of a public health emergency with unknown diseases and unknown clinical endpoints. Many HDBs will bring two-way channels of communication with researchers even to community-based, front-line points of care. This will make it possible to include swaths of the public as active research participants.

HDBs' contributions will mostly occur as infrastructure supplements in the operational conduct of trials (e.g., planning, recruitment, follow-up).[2] An entire $200 billion industry consisting of pharmaceutical companies, contract research organizations (CROs), academic centers, research organizations, and others exists to perform the highly specialized and rigorous work of conducting randomized, multicenter, endpoint-driven, double-blinded, placebo-controlled clinical trials. This work necessarily includes the management of trial data. That industry has already made great advances in meeting the challenges of multi-site, point-of-care, and virtual trials as are envisioned in this OSTP initiative. Here HDBs will play a critical supporting role.

While HDBs are not yet a force in the U.S. health industry, OSTP should anticipate the possibility of their rapid growth in the coming years and prepare for the significant improvements and unexpected innovations they can bring to the clinical research enterprise.

For example, HDBs offer secure patient-centered data repositories and exchange infrastructure that can be shared reliably by medical practitioners and medical researchers. This is unprecedented. It eliminates costs and efforts that would arise from assuming that the clinical research enterprise and the health care delivery system will remain fundamentally separate domains as they have been for the past 60 years.

That domain separation underlies the RFI's Use Case. There, the first priority in preparing to conduct trials is to identify and incentivize clinical trial sites and then, secondarily, to rely on those sites to recruit research participants. HDBs obviate that first step and turn the second on its head. HDBs would enable the federal public health authorities to reach out directly to patients (potential participants) who meet the required medical profile. Then, secondarily, trial

sites or, if applicable, patients' own points of care could be identified and enlisted in the research.

Depending on particular HDBs' business models and operating protocols, many patients with HDB accounts will have already pre-consented for observational study. They will already have at least some familiarity with medical research. This reduces the friction of consenting them as participants in interventional trials, and provides a strong cadre of well-documented patients to become involved in trials. HDBs' two-way communications paths enabling convenient re-contacting of clinical trial participants provide pragmatic flexibility as new biomarkers of interest are identified over the course of trials.

These infrastructure features allow significant savings in time and cost. They demonstrate the kind of benefits from using HDBs early in a crisis as essential components of emergency trials infrastructure.

HDBs could simplify and expedite organizing clinical trials in other practical ways, beyond what is envisioned in the RFI's Use Case. These capabilities arise from researchers' interactive engagement with patients and at points of care. These contacts can occur well before a trial protocol is developed, enabling it to be developed using FDA's Patient-Focused Drug Development Guidance. Even while researchers are still working on the new molecular targets in the laboratory, HDBs will enable advance work with the nation's front-line providers and patient population. This can make ready a large cohort of participants, with background medical history data already entered.

The broad range of value that HDBs would bring to the clinical trials infrastructure suggests that OSTP consider broadening the scope of these RFIs. The work of designing trials begins in the very early, urgent, and normalized collection and analytics of raw real-world data at the emergency's very outset. This collection can be accomplished more reliably and efficiently at scale from HDBs than from EHRs. Once public health authorities have a stable, initial picture of the disease, protocols for *observational* studies, such as might contribute data to registries, can be developed – also by central federal authorities – and sent out to patients. In this way a robust knowledge base can be built early, helping to avoid research silos from forming in the chaos of the emergency. HDBs also are a direct-to-consumer channel for collecting the patient voice on any aspect of the research at scale.

Yes, technically rigorous RCTs are necessary before vaccines or treatments can be approved. And yes, such trials will require appropriately trained and equipped sites and staff, and the ability to deploy in communities and populations across the nation using remote patient monitoring devices and the like. However, the clinical trials industry is already prepared to conduct such centralized, multi-site, clinical trials (ironically known as "decentralized clinical trials" – see Decentralized Trials & Research Alliance), with all necessary specialized components, such as visiting research nurses, remote patient monitoring, lab protocols, electronic case report forms (eCRFs), and so forth.

Because lack of participants is a far greater problem for clinical trialists than lack of sites, HDBs offer an ideal infrastructure for recruitment of research participants. After all, most HDBs' account service protocols will have systems to notify account holders of research opportunities of interest to them. Further, under some HDB business models, HDB account holders who want to participate in research may already have been pre-consented for observational research, smoothing meaningful informed consent for trials.

In addition, regulatory authorities increasingly require post-trial follow-up on participants and post-market follow up on patients. They will surely insist on these protocols in cases of emergency outbreaks or epidemics. This is among the infrastructure functions that HDBs are suited for. Indeed, public health emergencies will generate interest in long-term epidemiological studies that may go on for many years after crises wane. HDBs can both facilitate those research communication channels and bring pre-crisis medical histories of the study participants into the studies.

In short, HDBs introduce radical innovations and nuanced enhancements in the clinical research space. OSTP can expect them lead to significant improvements in the efficiency and evidence-generation power of clinical research. These features will be tested in pilots.

Directing Federal Funding to Accelerate FHIR Standardization, and Away From TEFCA

Paragraph 9 of the RFI asks how TEFCA could be used to support clinical trials, whether under currently authorized "Exchange Purposes," e.g., Public Health, or under "a future research-focused Exchange Purpose." The answer is: TEFCA (as ONC is implementing it) is wholly unsuited to support clinical trials now and in the future.

HDBs are the alternative bundle of infrastructure technologies to support clinical trials and medical research generally, and OSTP can expect HDBs to begin emerging this year. The reason is that January 1, 2023 was the deadline by which initial versions of standardized FHIR-based application programming interfaces (APIs), required for EHR certification under ONC's Interoperability Rule, were required to become widely available to consumers via their third-party application programming interfaces ("apps").

Recognizing the likely emergence of HDB infrastructure will help guide OSTP in directing government funding and attention toward the goals set forth in this RFI – health data quality and communication flows to enable a robust clinical trial infrastructure and data capture, in emergencies and otherwise.

As we explain below, OSTP should recommend to the President that the federal government redirect funding away from TEFCA and toward FHIR standardization. The goal is to speed standardizing and expanding FHIR APIs.

OSTP also should seek legislation to authorize a regulatory framework for the emerging Health Data Bank industry. A regulatory framework is essential to protect the public, earn its trust, and assure rapid expansion of HDB infrastructure for clinical trials and other care and research purposes.

Conserving federal resources for these purposes is imperative. OSTP should therefore seek Presidential directives to revise ONC's implementation of the Trusted Exchange Framework and Common Agreement (TEFCA) under the 21st Century Cures Act. The Cures Act's statutory *network-preservation mandate* is optimally implemented by encouraging replacement of the fax system – the de facto standard at present for exchanging health records – with secure, point-to-point digital pathways via HDBs. This will *preserve clinicians' and hospitals' networks*, endowing them with new flexibility in a patient-centered system architecture.

It is necessary to emphasize that the TEFCA mandate in Section 4003 of the Cures Act *does not specify preservation* of *any particular categories of health information networks, and does not specify that TEFCA preserve HIEs or networks of HIEs.* Rather, architecting TEFCA to preserve *hospital and medical office health information networks* is sufficient to meet the Cures Act's network preservation requirement for TEFCA.[3]

The President should therefore require ONC to discard the current, initial iteration of TEFCA and create a wholly new version of TEFCA. The re-conceptualized TEFCA should be based on, and support, nationwide, secure, point-to-point, FHIR-based, streamlined health data exchange. This is a drastic change; but it is necessary in order for TEFCA implementation to bolster, rather than detract from, other features of the Interoperability Rule and the patient-centric policy objectives enacted in the Cures Act.

Policy objectives and technological reality support HRBA's stark recommendation here. Designing TEFCA (as ONC has done) to attempt to preserve moribund Health Information Exchange (HIE) networks is ultimately futile. HIEs are unwieldy, institution-centered system designs. They are not patient-centric or even patient friendly. And they therefore will not be friendly to researchers, including the clinical trial community, seeking to streamline, strengthen, and accelerate clinical trial processes.

HIEs were created to try to help moving health data among providers, hospitals, insurance entities, and other institutions because routine, efficient, point-to-point, digital data transfer among disparate EHR systems was impossible. When HIEs were created, no national, digital, health information exchange standard existed to move data routinely stored in disparate, siloed EHR systems. *Now that the era of standardized FHIR-based data exchange is beginning, some HIEs may elect to convert to Health Data Banks. Other HIEs, now obsolete, will go out of business.*

ONC's current design for TEFCA under the Common Agreement requires TEFCA signatories – QHINs and downstream participants and sub-participants – to respond to nationwide query messages seeking data on particular patients. This discredited "shotgun query" or "record locator query" design cannot feasibly be implemented. It would overwhelm networks' capacities for throughput, create unsolvable patient matching problems with associated privacy rule violations, and so create cascading liability issues.

The idea of making record locator problems worse (as TEFCA does) by inserting regional brokers and disparate, local, voluntary exchanges only makes the system more costly and chaotic, and even more fraught with security and privacy issues. Access control and user authentication are well known problems that multiply at an accelerated rate with scale in such systems. There is no longer any reason to continue to support this inefficient, error-prone architecture.

Supporting clinical trial research, whether or not in emergency conditions, requires streamlined communication between researcher and patients for recruitment, ongoing protocol execution, and follow-up. These requirements cannot be met by TEFCA's SOP for

[3] The full text of Section 4003 of the Cures Act, titled "Interoperability," is available at https://www.healthit.gov/sites/default/files/facas/2018-02-23_TEF_TF_21stCenturyCures_4003_508.pdf

Individual Access Services.[4] It is beyond cumbersome, in practical effect a barrier to patients' use of TEFCA to obtain their complete medical records and communicate those records to their providers and to trialists and other researchers.

It is noteworthy how far the SOP for Individual Access Services departs from the Cures Acts *mandatory* specifications set out in reporting requirements for certification in sections 4002 of the Cures Act as it amends HITECH sections 3001(c)(5) and adds section 3009(a), as follows:

- EHR data must be exchangeable "*without special effort"* on the part of users. (Patients and physicians are among "users" under the Cures Act.) (HITECH as amended, new §3000(10)(A), as added by Cures Act §4003(a); emphasis added.)

- EHR data exchange must allow "*complete* access, exchange, and use of all electronically accessible health information for authorized use [under applicable law]." (HITECH as amended, new §3000(10)(B), as added by Cures Act §4003(a); emphasis added.)

- EHR data exchange cannot be implemented by ONC in ways that restrict "exporting *complete information sets*" as part of access to, or exchange of, health information. (HITECH new §3022(a)(2)(C)(i), as added by Cures Act §4004; emphasis added.) This means export of all of a patient's health records in the EHR system if a patient so requests.

- EHR data exchange must allow "access *to all data elements* of a patient's electronic health record" permitted by privacy laws. (HITECH new §3001(c)(5)(D)(iv) as added by Cures Act §4002; emphasis added.)

- EHR data exchange *cannot* be implemented by ONC in ways that *"are likely to substantially increase the complexity or burden"* of access to, or exchange of, health information. (HITECH new §3022(a)(2)(B) as added by Cures Act §4004; emphasis added. This provision perforce imposes a specific requirement for nationwide standardized exchange.)

- EHR data exchange must be enabled through the use of application programming interfaces or *successor technology or standards*. (HITECH new §3001(c)(5)(D)(iv) as added by Cures Act §4002; emphasis added. FHIR-based data exchanges and HDBs are successor technologies.)

- EHR data exchange must provide *the patient or an authorized designee* with a *complete copy* of his or her health information from an electronic record *in a computable format*. (HITECH new §3000(10)(B) as added by Cures Act §4003;

[4] TEFCA's numbingly elaborate Individual Access Services SOP is available at https://rce.sequoiaproject.org/wp-content/uploads/2022/09/Final-SOP-IAS-Exchange-Purpose-Implementation.pdf .

emphasis added.)

TEFCA's SOP for Individual Access Services fails all these specifications. It is a formidable barrier to convenient, comprehensive use by patients and consumers for health data exchange purposes in, among other settings, clinical trials.

To recapitulate: the Interoperability Rule supplies the first, FHIR-based iteration of a national, digital, health data exchange standard; and the standard will expand rapidly and infinitely with the Interoperability Rule's Standards Version Advancement Process. Federal funds should be spent on accelerating expansion of FHIR-based exchange standards under the Standards Version Advancement Process (SVAP). The goal is to support an ever-widening, and ever-deepening, scope of medical specialty data requirements and associated research protocols.

In these comments, we ask OSTP to recommend to the President that he intervene with the Secretary of HHS. The aim is to require ONC to reconceptualize and reformulate TEFCA. That is startling, because ONC has devoted time, significant federal funding, and commitment to implement an institution-centric vision of the Cures Act's TEFCA mandate. That includes designating an expensive "Recognized Coordinating Entity" to help develop and maintain TEFCA's Common Agreement for the benefit of HIEs.

But TEFCA, as ONC now is implementing it, is more than just a stark misuse of federal resources to sustain antiquated and unnecessary HIEs. *ONC's TEFCA architecture defies sound cybersecurity design principles. It envisions a complex mass of cobbled-together network nodes and pathways, and unwieldy, convoluted operating procedures that invite security penetration. The network architecture is suffused with specific vulnerabilities, all gratuitous weaknesses. This TEFCA cannot, for any practical purposes, be secured to a satisfactory level. It is the opposite of ONC's goal of "a universal floor for [trusted] interoperability across the country."*

TEFCA at present is therefore not a network architecture to rely on for biodefense or any other national security purpose, including, and not limited to, emergency clinical trials.

We emphasize that the network of networks and turgid operating procedures envisioned in ONC's current plans for TEFCA are also unsuited – inefficient past the point of being dangerous – for medical research purposes.

If forced to use this version of TEFCA, the clinical trial community – research institutions, clinical trialists, health care providers interested in clinical research, contract research organizations (CROs) and other clinical trial service providers, pharmaceutical and biotechnology companies, and community health care organizations – will face constant unnecessary recruitment obstacles, process delays, ongoing regulatory compliance problems, and constant undue expense.

Conclusion

Health Data Banks' emergence will create new, technologically advanced infrastructure that is available at all times for clinical trials – especially important in emergencies. The federal government should direct its funding, attention, and support to accelerating development of FHIR-based health data exchange to make HDBs ever more capable. Funding for more rapid

standardization of FHIR resources under the Implementation Rule's Standards Version Advancement Process is crucial to achieving these goals as quickly as possible.

In parallel, OSTP should recognize that ONC's implementation of TEFCA as it now stands is an obstacle to secure, effective nationwide health data exchange – where what is needed simply is FHIR-based, point-to-point data exchange to replace today's reliance on facsimile. TEFCA's insecure architecture is intolerable. To fulfill its responsibilities to the country, OSTP should begin the process of conceptualizing TEFCA anew, so it can be implemented in consonance with the Cures Act's basic policy goal of organizing health data around patients, not institutions or networks of institutions. This is a call for Presidential intervention.

Respectfully submitted,

The Health Record Banking Alliance

/s/ Richard D. Marks
Richard D. Marks, Vice President
richardmarks@earthlink.net

Request for Information (RFI) on Clinical Research Infrastructure and Emergency Clinical Trials:

Data Collection for Emergency Clinical Trials and Interoperability Pilot

Trials 87 FR 64821

Submitted by Gregory Hess, MD MSc MBA, Henrik Toft Sorensen MD PhD DrMedSci DrSc, John Acquavella PhD, Erzsébet Horváth-Puhó MSc PhD, and Sean Redmond MS

Represented Stakeholder Types: experienced clinical trialists, real-world researchers, health care providers interested in clinical research, contract research organizations (CROs), pharmaceutical and biotechnology companies, and community health care organizations.

Primary Objective: To ensure that coordinated large-scale clinical trials can be efficiently carried out across a range of institutions.

Secondary Objectives:

- To provide input on viable technical strategies for distributing clinical trial protocols and capturing clinical trial data, using common application programming interfaces (APIs).
- To assess the value of a pilot or demonstration project to operationalize data capture.

Executive Summary:

Starting with the end in mind to optimally enable emergency clinical trials and clinical research for public health, the proposed project is entirely feasible and will create significant value. Our experience and that of others illustrates the key components necessary for success: highly complete, accurate and timely data optimally gathered on a continuous basis in a national data repository specifically designed for this purpose. If properly designed and maintained, the repository can enable a wide variety of additional uses to support and progress public health.

Background:

Our assumption is that existing real-world patient-level data will be collected and used in clinical trial(s). The optimal data profile would be one source, in which data fields are:

- Complete, i.e., each data field is fully populated. Text notes would be of additional value. However, given the current state of optical character recognition, natural language processing, and other barriers, the cost/benefit likely does not justify their inclusion for the purposes of this RFI.
- Accurate, i.e., each data field has a value within pre-specified, valid parameters for the given element (*e.g.*, weight, BMI, etc.), which is correct for the patient.
- Timely, i.e., the data reflect the patient's recent status and healthcare encounters.

Electronic medical records (EMRs) are the most logical primary data source for emergency clinical trials. In theory, they include the vast majority of information needed to determine whether a patient is eligible to enroll in a clinical trial. EMRs include demographics, physiologic measures, diagnoses, procedures, medications, laboratory test results, radiology and imaging reports, and additional data available in add-on specialty modules. The practical challenges in their use for clinical trials include records that are incomplete, inaccurate, and potentially biased to enhance billing revenue. The 100+ brands of EMRs on the market show tremendous variation in data tables and layouts. Many EMRs lack significant quality assurance (QA) or quality checks for entered data. Patients' longitudinal medical history can be significantly

limited given the fragmented US healthcare system, with many sites of care, many providers, and frequent changes in payers.

The creation of fully integrated patient-level platforms or data warehouses combining EMRs with other data sources such as administrative and medical registries, administrative claims data, and adjunct data could significantly improve data completeness and quality. Such platforms/data warehouses can create a more comprehensive electronic health record (EHR) for a wide range of uses, including rapid eligibility screening for clinical trials, estimation of sample size/power, trial enrollment, design, implementation, and outcomes assessment. In a setting of an emergency clinical trial presumably the effect size of a vaccine or other intervention would be substantial, and a relatively modest scale could be sufficient. With incremental budget and ambition, a more comprehensive size and scope would enable a wider range of additional public health uses beyond an emergency, clinical trial situation.

In addition to data quality and completeness, a significant but achievable requirement is access to near real-time data. There are a range of strategies to consider – each with inherent strengths and limitations. These include, but are not limited to:

- Use of Sentinel Sites. This strategy relies on sampling to represent the 'whole' picture, incorporating such representative characteristics as patients´ age, gender, location, comorbidities, and exposures. Sentinel sites are used by the Centers for Disease Control (CDC), National Institutes of Health (NIH), and the National Oceanic Atmospheric Administration (NOAA) for tracking incidence and prevalence and making projections.

This approach reduces the amount of data collection and number of data sites to manage, while providing good population estimates and specific cases (patients) for clinical trials. After a successful launch, a 'pilot' project can be expanded to more sites, up to the entirety of the US population. Sites could include all types of healthcare providers, including community practices, pharmacies, and tertiary hospitals.

- Distributed Partners. This strategy represents a variation on sentinel sites. It has been used on a relatively large scale for retrospective research, including the Observational Medical Outcomes Partnership (OMOP) funded by the Foundation for the National Institute of Health (NIH). In brief, data in its native format at each OMOP site and source were transformed into a common data layout to identify the impact of variations in data sources and curation on outcomes, as well as optimal statistical methods for analyses. On an ongoing basis, a series of sufficiently large and representative distributed partners could be contracted to routinely extract data and populate a common central data repository, or provide automated access to their distributed data.
- Licensing from Commercial Healthcare Data Warehouses. A number of commercial enterprises in the US currently gather regional and/or national healthcare patient-level data on a daily basis. This approach also greatly reduces the number of data sites to manage and its low lag time provides for very recent data. However, the approach is relatively costly and often lacks transparency regarding data limitations and transformations resulting from 'homogenizing' hundreds of primary sources.
- Dominant EMR Licensing. Given the limitations noted above for the range of existing EMRs, an approach based on the Pareto Principle would entail working with one

manufacturer initially. For example, EPIC is reported to have the largest market share of acute care hospital installations (≈ 30%). Rights to license EPIC data would be required. Advantages of this approach are reduced management burden, with one point of contact, and a clear understanding of data provenance. Over time, additional EMR manufacturers potentially could be enrolled in the program.

- Creation of a National Data Repository for Emergency Clinical Trials. This approach is optimal for creating data expressly 'fit for purpose' for emergency clinical trials. Rather than starting from the supply-side by simply repurposing existing data and systems, this market-driven demand design "starts with the end in mind." Importantly, it *may* use and leverage existing data and systems, but it is not constrained by them. (Details are beyond the scope of this RFI.) Relevant US case examples include nationwide disease-based, integrated patient-level platforms expressly designed for cardiology (*e.g.*, CLIPPER, Hess et al AHA QCOR 2015), oncology (*e.g.*, OSCER, Hess et al ASCO Annual Meeting 2008 Jrnl Clin Onc), and others. The Danish system of registries is an excellent example in which 'the entire country is the cohort (study population),' making possible rapid retrospective studies and prospective trials. While this approach may have a higher short-term direct cost, it has a lower long-term total cost, high value, and the highest probability of success in meeting goals.

Based on the Use Case presented in the Supplementary Information, a number of steps can be eliminated or truncated including:

1) "The eCRFs would be transmitted electronically via common APIs to the

sponsor."

2) "A study site's health IT system would present the eCRF content to clinicians in a manner that expedites data collection and (ideally) fits within clinician workflows."
3) "As the clinician obtains data elements to complete the eCRF, the data would be captured in the patient's electronic health record."

Continuous direct data collection from sites could eliminate these steps and would allow the sponsor's dedicated staff to populate the eCRF in a reliable, timely manner without requiring an on-site clinician to obtain data elements. In terms of consent, the Health Insurance Portability and Accountability Act (HIPAA) states that the *"... Privacy Rule expressly permits disclosures without individual authorization to public health authorities authorized by law to collect or receive the information for the purpose of preventing or controlling disease, injury, or disability, including but not limited to public health surveillance, investigation, and intervention.* Following screening of the subset of patients identified in this centralized and automated process, participating clinicians would perform validation and procure informed eConsents at the care site.

While technical transfer of data is a significant component for success, key considerations are ensuring that the data fields are complete and accurate and have little lag time. Our extensive experience suggests that if the goal is nearly complete data capture across a broad range of healthcare sites and types, then multiple options must be available. These include (but are not limited to) on-line e-forms, executable (.exe) files for batch exports, and use of application interfaces (APIs). Alternatively, if a representative sample is preferred, then generally fewer data transfer options are required for success.

Given the many comparable projects that have been implemented successfully for nationwide data collection, surveillance, and rapid clinical trials, a pilot or demonstration project is not required *per se*. However, the project typically would benefit from beta testing prior to full deployment. User validation and acceptance with a feedback loop and a live help desk also are important components. In addition, planning should include compiling a list of 'bugs,' fixes', and desired features from end users, with a systematic plan for continuous product improvement.

Limitations:

Please note that the recommendations and information we have provided are of course shaped by a range of assumptions at both a macro level for an envisioned emergency clinical trial and a more micro level such as effect size, outcomes, endpoints and other considerations.

Conclusion:

In conclusion, we strongly support and endorse the Office of Science, Technology and Policy's goal of Data Collection for Emergency Clinical Trials. It is entirely feasible with nearly certain success, if well-designed and implemented. From every perspective, including cost/benefit and other economic analyses, as well as clinical and humanistic analyses (ECHO model, Kozma *et al.*), the initiative can create tremendous value for enhancing US public health and preparedness.