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.