

A framework for representing clinical research in FHIR

Hugo Leroux^{1,5}[0000-0002-2033-8178], Christine K Denney^{2,5}[0000-0002-7180-6867], Smita Hastak^{3,5}[], and Hugh Glover^{4,5}[]

¹ The Australian E-Health Research Centre, CSIRO, Brisbane QLD 4029, Australia
hugo.leroux@csiro.au

² Eli Lilly and Company, Lilly Corporate Center, Indianapolis IN 46285, U.S.A.
christi_d@lilly.com

³ Samvit Solutions, Reston VA 20190, U.S.A.
shastak@samvit-solutions.com

⁴ Blue Wave Informatics LLP, Exeter, EX4 5AH, United Kingdom
hugh_glover@bluewaveinformatics.co.uk

⁵ HL7 Biomedical Research and Regulations Working Group
<https://confluence.hl7.org/display/BRR/>

Abstract. The benefits of clinical research have been widely acknowledged. However, clinical research is often costly, time-consuming, and burdensome to both the participants and researchers. There has recently been much emphasis on the need to streamline how clinical research is conducted and maximise the benefits of research through the sharing of research data and methods. In this paper, we explore the suitability of the Health Level 7 FHIR standard for representing and managing clinical research. While FHIR has gained popularity within patient care, the development of FHIR models and solutions to facilitate the delivery of clinical research is still in the early stages of maturity. This work outlines the activities of the HL7 Biomedical Research and Regulations FHIR working group in developing FHIR-based models and solutions for designing and conducting clinical research more effectively. Our goal is to ascertain whether a native, FHIR-based, API definition is suitable for clinical research, can alleviate the issues relating to both the discoverability and accessibility of clinical research data, and enable the semantic interoperability of the data that can lead to the reusability of the datasets. We outline how the FHIR resources have the potential to overcome the challenges of sharing and reusing clinical research data. We discuss some of the current limitations associated with those resources and how we are working to address them. Our overarching goal for this work is to stimulate a *robust* discussion on how clinical research semantics and data exchange use cases could be represented in FHIR.

Keywords: Clinical Research · Data Sharing · FHIR · Data model · FAIR.

1 Introduction

Clinical research is an integral and important part of healthcare delivery. A report on the economic benefit of clinical research data sharing in Australia has found that the Australian government invests \$1.5 billion in health research and development annually [1]. It estimates that the value to the Australian Gross Domestic Product could exceed \$129 million annually if data from publicly-funded clinical research was made accessible to the research community [1]. However, the effectiveness of clinical research relies on its ability to have an impact on health [2].

Clinical research is costly, time-consuming and taxing both on the clinical researchers and on the participants. There has been much emphasis, lately, on the need to streamline the way in which clinical research is conducted to maximise the benefit [2, 3]. There has also been a push for clinical research data and methods to be shared more broadly. Warren [4] stated that *‘data sharing may help reduce costs by allowing researchers to avoid duplicating trials or to answer questions without undertaking a separate data collection effort’*.

A common theme across data sharing initiatives is the *‘idea of building infrastructures based on rich metadata’* that will *‘support their optimal re-use’* [5]. Mons et al. [5] stated that ensuring that all resources are findable, accessible, interoperable and reusable (FAIR), *‘requires widely shared and adopted standards and principles’*. FAIR refers to a set of principles focused on ensuring that research objects are reusable, will be leveraged, and become as valuable as possible [5]. There has been an increasing focus on reproducibility and replicability of clinical research [2] resulting from findings that over 70% of published research cannot be reproduced by others [6].

The *Fast Healthcare Interoperability Resources* (FHIR) framework is an emerging standard that is geared towards the communication of clinical data using HL7 messaging protocols and, when supported by a rich information model, can achieve the *semantic interoperability* of clinical data. As FHIR is gaining importance within the healthcare and life sciences community [7] and has been swiftly adopted by the major healthcare providers (including *Cerner* [8] and *Epic* [9]), FHIR is likely to play a significant role in the future of healthcare and clinical research [10]. Furthermore, the National Institutes of Health have issued a notice to *‘explore the use of the FHIR standard to capture, integrate, and exchange clinical data for research purposes and to enhance capabilities to share research data’* [11]. The current effort in FHIR resource development is primarily focussed on patient care and geared towards electronic health records (EHRs) and hospital billing and accountancy systems. Developing FHIR models and solutions to facilitate the delivery of clinical research is still in the early stages of maturity, notwithstanding some early efforts [12, 13].

Our overarching goal in this project is to ascertain whether a native, FHIR-based, data model is suitable for clinical research, can alleviate the issues relating to both the discoverability and accessibility of clinical research data and, enable the semantic interoperability of the data that can lead to the reusability of the data sets. In addition, we believe that the adoption of the FHIR standard for

developing clinical research protocols and capturing clinical research data can also help preserve the integrity of the data and the privacy of individuals through the adoption of profiling to constrain the content exposed by the resource.

In the next section, we elaborate on the considerations for representing clinical research in FHIR. We outline the activities of the HL7 Biomedical Research and Regulations (BR&R) FHIR working group (WG) in developing FHIR-based resources for designing and conducting clinical research more effectively. We discuss how the FHIR resources have the potential to overcome the challenges of sharing and reusing clinical research. We then discuss some of the current limitations associated with the FHIR resources and how we are working to addressing them.

2 Representing Clinical Research in FHIR

The core components in FHIR are *resources*, which are logical constructs in healthcare and define both behaviour and meaning. The resources are scoped to the most commonly known data exchange implementation needs and collectively form and support the complex health systems. Extensions are a mechanism provided by FHIR to allow support for the less common or outlier use cases of data exchange whose requirements are not in the scope of the base resource definition. To ensure interoperability, FHIR also enables the creation of profiles that can be used to constrain the structure of the resource, using some rules defined by the profiler, to ensure compliance by the implementation systems. The standardisation of the methods is achieved by defining a set of common functionality within the resources while the standardisation of the data semantics is facilitated by allowing and occasionally enforcing the definition of code systems and value sets that describe the data.

The HL7 BR&R FHIR WG has been established to facilitate the development of common standards and the management of research-focussed domain analysis models for clinical research information management. BR&R also seeks to assure that related or supportive standards produced by other HL7 groups are robust enough to accommodate their use in regulated clinical research. A shared semantic view is essential if the clinical research community is to achieve computable semantic interoperability. In this regard, the BR&R and Clinical Decision Support FHIR WGs have developed a small number of resources (namely `ResearchSubject`, `ResearchStudy`, `PlanDefinition` and `ActivityDefinition`) for describing clinical research study design in FHIR. These four resources are still in early stages of design and therefore are at low levels of maturity. The data are expected to be captured using existing FHIR resources such as `Encounter`, `Procedure` and `Observation` to name just three. This is anticipated to expedite the sharing of clinical research data in the future.

Sharing of clinical research data has numerous challenges relating to the discoverability of the clinical research undertaken, the availability of the data sets and associated methods in a machine-readable, structured and standardised manner, and the adoption of common standards. Addressing these challenges

could produce results and methods that are more easily understandable and facilitate the reproducibility and replicability of the results. We introduce the aforementioned resources below and elaborate on how they address the challenges associated with sharing clinical research data.

2.1 FHIR Resources for Clinical Research

ResearchStudy. The `ResearchStudy` resource provides a template for the definition of the overall structure of a study or trial, including the protocol and the various *arms* comprising the study. It provides references to the `PlanDefinition` resource to allow the user to define the protocol for the study; to the `Organization` resource to define the sponsor; to a `Practitioner` resource to define the principal investigator; and to a `Location` resource to facilitate the description of a study's site physical property. Other study characteristics, such as the study identifier, the title, the description, and the category of study can be defined within the core resource.

ResearchSubject. The `ResearchSubject` resource facilitates the definition of a participant to the study. It provides two mandatory references: one to the `ResearchStudy` and the other to the `Patient` resource. The purpose of the latter is to link an actual patient to the role of participant in the study. Furthermore, it provides a reference to the `Consent` resource to facilitate the participant consenting to participate in the study.

PlanDefinition. A `PlanDefinition` is a pre-defined group of actions to be taken in particular circumstances, often including conditional elements, options, and other decision points. The resource is flexible enough to be used to represent a variety of workflows, as well as clinical decision support and quality improvement assets, including order sets, protocols, and decision support rules. Although this resource currently does not fully support the clinical research use cases, it has a good foundation to be leveraged for defining the protocol in relation to the complex schedule of activities, objectives, and outcomes. HL7 BR&R WG members are currently evaluating this resource to identify and map the protocol concepts, identify gaps, provide updates to definitions, and possibly consider developing extensions and eventually a Clinical Research or Protocol FHIR Profile.

ActivityDefinition. An `ActivityDefinition` is a shareable, consumable description of some activity to be performed. It may be used to specify actions to be taken as part of a workflow, order set, or protocol, or it may be used independently as part of a catalog of activities, such as orderables. Within clinical research, this resource would define all the activities that are defined in a protocol. This may include administrative activities such as checking eligibility, trial enrolment, obtaining consent, and capturing the various clinical activities such as blood collection, urine analysis, etc.

2.2 Information Model

Figure 1 illustrates a set (or network) of HL7 FHIR Resources and their relationships that are relevant to a clinical research use case of a **Patient** in a role of **ResearchSubject** who is a participant in a **ResearchStudy** that is sponsored by an **Organization**, being conducted at a particular **Location**. The **ResearchStudy** is being executed based on the protocol definition in a **PlanDefinition**, which includes **ActivityDefinitions**. The **ActivityDefinition** describes a **CarePlan** for each participant that further defines **Appointments**, which lead to **Encounters** that produce **Observations** that relate to a particular patient within the study.

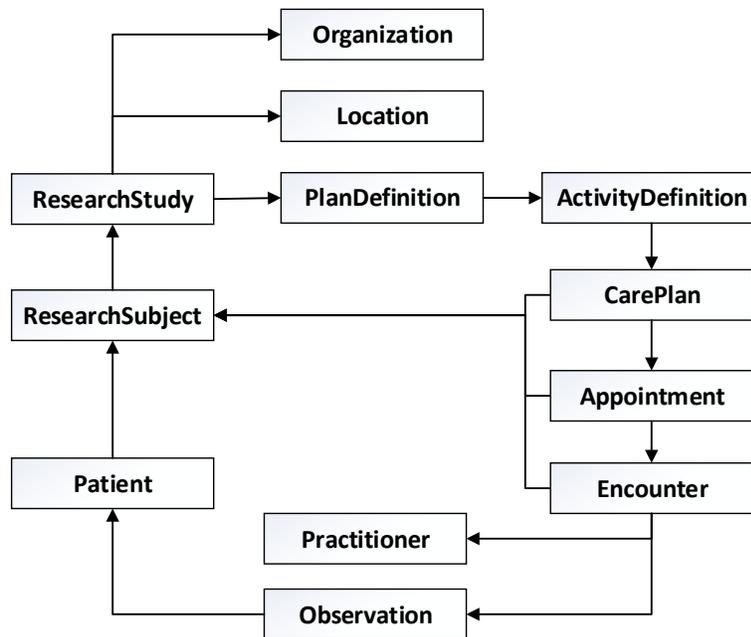


Fig. 1. FHIR Information Model for Clinical Research.

3 Overcoming the Challenges of Sharing and Reuse

We believe that the information model described previously should overcome the challenges of sharing and reusing clinical research data. The BR&R WG, along with other working groups, have engaged in a number of initiatives to promote clinical research data sharing and reuse. We describe two related focus areas below.

3.1 Activities within BR&R in promoting data sharing and reuse

BRIDG Mapping. The Biomedical Research Integrated Domain Group (BRIDG) Model [14] is a collaborative effort engaging stakeholders from the Clinical Data Interchange Standards Consortium (CDISC), the HL7 BRIDG Work Group, the International Organization for Standardization (ISO), the US National Cancer Institute (NCI), and the US Food and Drug Administration (FDA). The goal of the BRIDG Model is to produce a shared view of the dynamic and static semantics for the domain of basic, pre-clinical, clinical, and translational research and its associated regulatory artefacts.

The BRIDG model is supported by the HL7 BR&R WG as its domain information model and is intended to provide the semantic foundation to the artefacts developed by BR&R. It is a conceptual model, although parts of the model are quite granular and therefore often considered a hybrid of conceptual and logical layers. BR&R WG members are leveraging the BRIDG model concepts, definitions, and relationships to inform FHIR resource models.

CDISC Lab Semantics in FHIR. The BR&R WG and the Orders and Observations (O&O) WG cosponsored the development of an implementation guide [15] to provide direction for sites and sponsors seeking to exchange laboratory data via FHIR (Note: scope is limited to the data collected to evaluate safety of an interventional study medication).

3.2 The Availability of Machine-Readable Clinical Research Definition

The BR&R WG is exploring ways to make the clinical research study protocol available in a machine-readable manner. A structured and computable protocol is important for clinical research, yet the challenge has not been fully addressed by prior initiatives. The CDISC Protocol Representational Model (PRM) [16] is a UML-based standard that developed a set of standard protocol concepts and was intended to be used alongside the other CDISC and HL7 standards. PRM has now been integrated within the CDISC Controlled Terminology Package (CT) [17]. The main drawback of the CDISC PRM and CT standards is that they do not adhere to commonly used clinical terminology standards, such as SNOMED CT or LOINC, which makes semantic interoperability of the protocol difficult to achieve. Furthermore, the CDISC PRM has had limited adoption by the clinical research community [18]. Another initiative, SPIRIT [19], provides a checklist for a 33-item trial protocol to be entered electronically. SPIRIT currently does not allow for coded input and only allows the protocol to be entered in free-text. Furthermore, it does not allow the protocol to be linked to either a controlled clinical vocabulary, such as SNOMED CT, nor to any publications discussing the study.

The desired future state would draw upon the existing initiatives and design a standardised, structured and computable representation of the study protocol

as a set of FHIR artefacts (resources, profiles, extensions) that define the approved protocol. This should enable the validation of the study data against the clinical research questions defined within the protocol and what was scheduled and performed during the study.

3.3 Adopting a Common Standard

There have been a number of initiatives lately to standardise the data for use both in healthcare and clinical research. The HL7 WG on Semantic Interoperability [20] is engaged in developing models and use cases in facilitating the use of RDF as a common semantic foundation for healthcare information interoperability. One of their key deliverables has been the development of the FHIR RDF representation and ontology [21]. FHIR RDF might prove useful for implementing our model due to the complexity of representing the bi-directional nature of clinical research. Another challenge is the gap between patient care and clinical research data standards. Aerts [22] hints at the convergence of the CDISC and HL7 standards to bridge that gap. Furthermore, there is a global effort to standardise clinical research data [2] to translate it into meaningful discovery and improve benefits to patients. Indeed, the *CDISC Lab semantics in FHIR* project [15] suggests that there is currently no standard in place to provide data to sponsors as they adhere to data standards within the CDISC suite of standards, whereas healthcare is progressively adopting the FHIR standard for communication and distribution [23, 24]. The need to standardise data is vital if we want to achieve meaningful use and semantic interoperability of clinical research data [12]

4 Discussion

Developing a framework for representing clinical research in FHIR is challenging but provides an important opportunity for change. We have the potential and responsibility to guide the next generation of clinical research through our engagement with researchers, sponsors, regulatory agencies, and industry. We present our thoughts below in helping to shape this important engagement.

4.1 Resource Context and Workflow

In many traditional domain models [14] for clinical research, the entities may be used in varying contexts and change state over time. For example, the visit concept may represent both a planned activity and the resulting performed activity. Over time, the attributes and status of the entity adapt to the process. Conversely, while not a limitation, FHIR differentiates resources by their role in the workflow process and provides separate resources for the template of the visit (`ActivityDefinition`), the designated template of the visit for a particular participant (`CarePlan`), the scheduled visit (`Appointment`), and the occurred

visit (**Encounter**). In order to correctly identify the desired resource, the implementer must understand the intended use of the resource and how to traverse the workflow in which it resides.

4.2 Adherence to an information model or foundation ontology

Resources in FHIR are built in a pragmatic manner to facilitate their rapid implementation. Consequently, the aforementioned resources are not based on a foundation ontology, such as the **Semanticscience Integrated Ontology** (SIO) [25] or the **Ontology for Biomedical Investigations** (OBI) [26]. It is also a common misconception that the FHIR resources equate to an information model. An information model such as BRIDG or CDISC PRM provides the concepts, meaning, and relationships between the concepts of a given domain of interest. These models can be used to inform the design of implementation-oriented FHIR artefacts. Our model, however, could benefit from some judicious mapping to the SIO and OBI ontologies. We seek to engage with the *Semantic Web and Life Sciences* community to help us facilitate this mapping.

4.3 Linkages to Clinical Research Resources

Clinical research, generally, cannot have visibility to the **Patient**, only to the **ResearchSubject**, to support de-identification and privacy of participants. However, creation of a **ResearchSubject** instance, in particular, requires a reference to the **Patient** resource. Consequently, a *dummy Patient* resource needs to be created to play the role of participant in the study. Furthermore, many of the FHIR resources, such as **Observations**, **Procedures** and **Diagnosis**, provide a *mandatory* reference to a **Patient** but not to the **ResearchSubject**. While not technically a limitation, it adds another level of complexity for traversing the model. The CDISC Lab Semantics in FHIR Implementation Guide provides some guidance to the sites on how to mask the patient identity.

4.4 Model Maturity

In assessing the maturity of the FHIR resources for use in clinical research, we see potential for enhancements to the currently defined **ResearchStudy** and **ResearchSubject** resources from BR&R as well as to many other resources that were defined with only clinical use cases - resources such as **Observation**, **Procedure**, etc. At present, the **ResearchStudy** resource contains attributes designed to capture a text description of the arms for the study. However, this information is defined during protocol development and therefore, it may be better to design the concept of *arm* as part of the **PlanDefinition** resource and remove it from the **ResearchStudy** resource. BR&R is currently discussing this change on the **ResearchStudy** resource. **ResearchStudy** also links to a **Location** to represent the study site overseeing a set of **ResearchSubjects**. However, this falls short of representing the full context of the study site, such as the site personnel and study participants assigned to a site.

4.5 Traversing the Model

By their very nature, FHIR resources introduce complex data types and relationships. This, coupled with the adoption of the Representational State Transfer (REST) framework, means that traversing a network of resources, as depicted in Figure 1, necessitates moving beyond the lens of a traditional relational design, which starts at a point and moves in a single direction, to looking at the model as an ontology of nodes. One such example is collating the participants in a research study. The `ResearchStudy` resource does not contain a reference to the `ResearchSubject` enrolled in the study. Rather, the reference is contained within the `ResearchSubject` resource. Similarly, when trying to elucidate an `Observation` related to a `ResearchStudy`, one has to traverse the `ResearchStudy` to obtain the relevant context, then one needs to work one's way back from the `Observation` to the `Encounter` to ultimately link the relevant observation to the `ResearchSubject` via the `Patient` resource.

5 Conclusion

There is an increasing need to streamline how clinical research is conducted and maximise the benefits of research through sharing of research data and methods. This work has explored the suitability of the HL7 FHIR standard to represent and manage clinical research. We have outlined the activities of the HL7 Biomedical Research & Regulations working group in developing FHIR-based models and solutions to design and conduct clinical research more effectively. We have proposed an information model comprising the FHIR resources to semantically represent the clinical research lifecycle, so as to facilitate semantic interoperability and increased sharing of the data. There have been a number of distinct standards proposed recently for representing clinical research. Our goal, for this work, is to stimulate a robust discussion on how clinical research semantics and data exchange use cases can be represented in FHIR.

References

1. CIE: The public benefit of collaborative access to publicly funded clinical and health studies (2019), <https://discovery.csiro.au/permalink/f/12s7o4e/CSIRO1196669610001981>
2. Jauregui, B., Hudson, L.D., Becnel, L.B., et al.: Global standardization of clinical research data. *Applied Clinical Trials* **28**(4), 18–24 (2019)
3. Kush, R.D., Nordo, A.H.: *Data Sharing and Reuse of Health Data for Research*, pp. 379–401. Springer International Publishing (2019)
4. Warren, E.: Strengthening research through data sharing. *New England Journal of Medicine* **375**(5), 401–403 (2016)
5. Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L.O.B., Wilkinson, M.D.: Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* **37**(1), 49–56 (2017)

6. Baker, M.: 1,500 scientists lift the lid on reproducibility. *Nature News* **533**(7604), 452 (2016)
7. HL7 FHIR: Argonaut project (2018), <http://argonautwiki.hl7.org>
8. Miliard, M.: Cerner touts adoption of normative fhir r4 standard (2018), <https://www.healthcareitnews.com/news/cerner-touts-adoption-normative-fhir-r4-standard>
9. Epic: Open epic (2018), <https://open.epic.com/Interface/FHIR>
10. Posnack, S., Barker, W.: Heat wave: The u.s. is poised to catch fhir in 2019 (2018), <https://www.healthit.gov/buzz-blog/interoperability/heat-wave-the-u-s-is-poised-to-catch-fhir-in-2019>
11. NIH: Fast healthcare interoperability resources (fhir®) standard (2019), <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-122.html>
12. Leroux, H., Metke-Jimenez, A., Lawley, M.J.: Towards achieving semantic interoperability of clinical study data with FHIR. *Journal of biomedical semantics* **8**(1), 41 (2017)
13. Leroux, H., Metke-Jimenez, A., Lawley, M.J.: ODM on FHIR: Towards achieving semantic interoperability of clinical study data. In: SWAT4LS. CEUR (2015)
14. NCI: About BRIDG model (2016), <https://bridgmodel.nci.nih.gov/about-bridg>
15. HL7 BR&R: CDISC Lab Semantics in FHIR Implementation Guide (2019), <http://hl7.org/fhir/uv/cdisc-lab/2019Sep/>
16. CDISC: Protocol Representation Model (2010), <http://www.cdisc.org/protocol>
17. CDISC: Controlled terminology (2019), <https://www.cdisc.org/standards/terminology>
18. Huser, V., Sastry, C., Breymaier, M., Idriss, A., Cimino, J.J.: Standardizing data exchange for clinical research protocols and case report forms: An assessment of the suitability of the Clinical Data Interchange Standards Consortium Operational Data Model. *Journal of Biomedical Informatics* **57**, 88–99 (2015)
19. Chan, A.W., Tetzlaff, J.M., Altman, D.G., Laupacis, A., et al.: SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Annals of Internal Medicine* **158**(3), 200–207 (2013)
20. HL7: RDF for Semantic Interoperability (2016), http://wiki.hl7.org/index.php?title=RDF_for_Semantic_Interoperability
21. Solbrig, H., Prud'hommeaux, E., Jiang, G.: Blending FHIR RDF and OWL. In: SWAT4LS. vol. 2042. CEUR (2017)
22. Aerts, J.: Towards a single data exchange standard for use in healthcare and in clinical research. *Studies in health technology and informatics* **248**, 55–63 (2018)
23. Siwicki, B.: How FHIR 4 will drive interoperability progress in healthcare (April 2019), <https://www.healthcareitnews.com/news/how-fhir-4-will-drive-interoperability-progress-healthcare>
24. Borfitz, D.: Imagining a world on FHIR (2019), <https://www.clinicalinformaticsnews.com/2019/05/02/imagining-a-world-on-fhir.aspx>
25. Dumontier, M., Baker, C.J., Baran, J., et al.: The semanticscience integrated ontology for biomedical research and knowledge discovery. *Journal of biomedical semantics* **5**(1), 14 (2014)
26. Bandrowski, A., Brinkman, R., Brochhausen, M., et al.: The ontology for biomedical investigations. *PloS one* **11**(4) (2016)