

ODM on FHIR: Towards Achieving Semantic Interoperability of Clinical Study Data

Hugo Leroux, Alejandro Metke-Jimenez and Michael J. Lawley

The Australian E-Health Research Centre, Health and Biosecurity, CSIRO
{firstname.lastname}@csiro.au

Abstract. Observational clinical studies play a pivotal role in advancing medical knowledge and patient healthcare. However, to lessen the prohibitive costs of conducting these studies and support evidence-based medicine, results emanating from these studies need to be shared and compared with one another. This paper explores how semantic interoperability of clinical data can be achieved by integrating two prominent standards for clinical data: ODM and FHIR. ODM lacks a rich-enough information model to adequately capture the contextual information of clinical study data. This is overcome by using FHIR's information model to achieve *semantic interoperability* of clinical data. This work outlines our ongoing effort to integrate the ODM standard to the FHIR standard. In particular, it demonstrates how the hierarchical ODM model lends itself to be mapped to the ubiquitous FHIR resources. We describe the approach and provide insights into the assumptions made to fit the clinical data extracted from the ODM standard into the FHIR resources. Our focus is not only on mapping the data from ODM to the FHIR models but on capturing the contextual information, present in other sources, such as the *study protocol*, and which should have been made available with the extracted data. Finally, we discuss the exceptions under which the extracted ODM data does not adequately fit the targeted FHIR resources and offer some insight into a suitable solution.

1 Introduction

It is increasingly important to share results from clinical studies. The challenge when comparing results from clinical studies is to ensure that we compare corresponding data sets. The CDISC ODM¹ defines an XML-based standard that has been mandated by the Food and Drug Administration (FDA) for the electronic capture and reporting of clinical study data. While ODM provides a vehicle to communicate the results back to the regulatory body, it lacks a rich-enough information model to capture the innate contextual information of the clinical study data [11]. As a result, ODM is ill-suited for advancing the *semantic interoperability* solution that is required to achieve cross-study exploration of the clinical studies.

¹ Clinical Data Interchange Standards Consortium Operational Data Model

A recent report [12] has advocated the integration of clinical data from disparate studies in a *machine readable* format to promote *efficient evidence-based medicines*. By the same token, Hsu et al. [8] argue that to fulfil precision medicine requires the mining and aggregation of clinical data from multiple sources and novel approaches to obtaining contextual observations. The Fast Healthcare Interoperable Resources (FHIR) framework, which is a HL7 standard that has been swiftly adopted by the health-care community, looks the likely candidate for overcoming this challenge. It is geared towards communication of clinical data using HL7 messaging protocols but is also supported by a rich information model to achieve semantic interoperability of clinical data. This makes FHIR the natural match to complement the ODM standard.

In this work, we present an approach to integrate the CDISC ODM standard with the FHIR resources to enrich longitudinal clinical study data extracted from ODM. The aim is to exploit the rich information model from FHIR to reintroduce the contextual information that is often contained within the *study protocol* documents but not contained within ODM and rarely made available alongside the clinical data. We explore the hierarchical concepts within the ODM and describe how these fit the pervasive FHIR resources. In so doing, we elaborate on the assumptions that have been made to adapt the extracted clinical data to the FHIR resources. The expected benefit of this approach is in facilitating a richer exploration and querying of clinical data coupled with relevant contextual information.

2 Background

2.1 CDISC ODM

The CDISC ODM data model [4] is specifically designed for a data capture context. It consists of two main hierarchies: a **Clinical Data** and a **Metadata** hierarchy, as depicted in Figure 1, that are referenced using the same **object identifier** (OID). These two parallel hierarchies ensure that the clinical study follows a predetermined structure of **subject**, **event**, **form**, **item group** and **item**. An **item** corresponds to a single measurement or analysis result captured during the study. An **item group** typically comprises a set of contextually-related measurements or results. A **form** (or case report form) is a collection of items, some grouped, for capturing and displaying clinical data. A **study event** corresponds to a patient encounter in the course of the study whereby data corresponding to one or more forms is collected.

2.2 FHIR

FHIR², published as a **Draft Standard for Trial Use** by the Health Level Seven (HL7) organisation is a specification designed to facilitate the exchange of

² <http://hl7.org/fhir/2015Sep/overview.html>

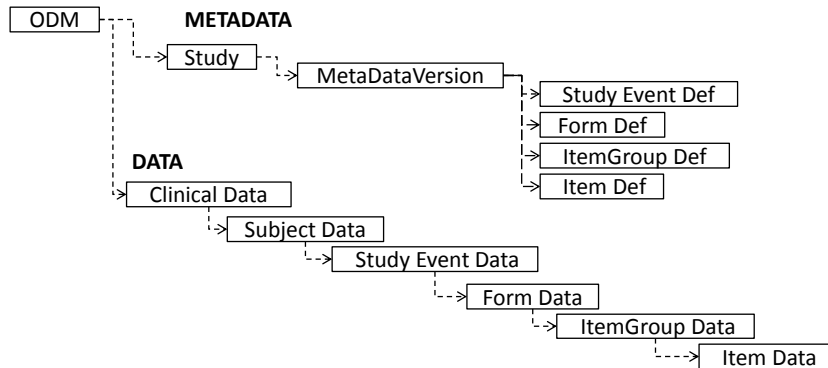


Fig. 1. Illustrates how the ODM data is organised into *data* and *metadata*

healthcare-related information. FHIR revolves around *resources*, which are snippets of *highly-focussed* data. The specification defines a set of minimalistic and generic elements and *extensions* are used to bridge the gap for the remaining content. Figure 2 depicts the hierarchy for the proposed FHIR resources to model the ODM data. The entities in red (CarePlan and Questionnaire) denote metadata concepts. The remaining entities model the clinical data at various levels. Solid lines are used to denote the links between the entities. The original model is depicted in Figure 4 in the Appendix. It was discarded in favour of Figure 2 so that there could be a clear demarcation between the *metadata* and the *data* and to allow the questionnaire to be linked to the care plan. The original model did not allow for a resource to contain the entire clinical data pertaining to a patient. This role is fulfilled by the `ClinicalImpression` resource in the new model and has necessitated the introduction of the `EpisodeOfCare` resource.

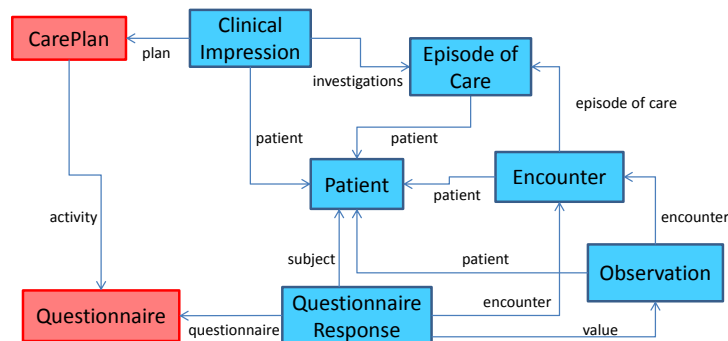


Fig. 2. Depicts the *metadata* (red) and *data* (blue) FHIR resources and their links

3 Integrating ODM with FHIR

This section describes the approach to integrate ODM with FHIR. We describe the components from the ODM hierarchy that we seek to map to FHIR. We assume here that the person doing the mapping has access to all the contextual information relating to the study. In this vein, it is not our aim to perform a one-to-one mapping of the CDISC ODM with FHIR. Rather we seek a holistic approach to mapping the hierarchical concepts between the two models as depicted in Figure 3.

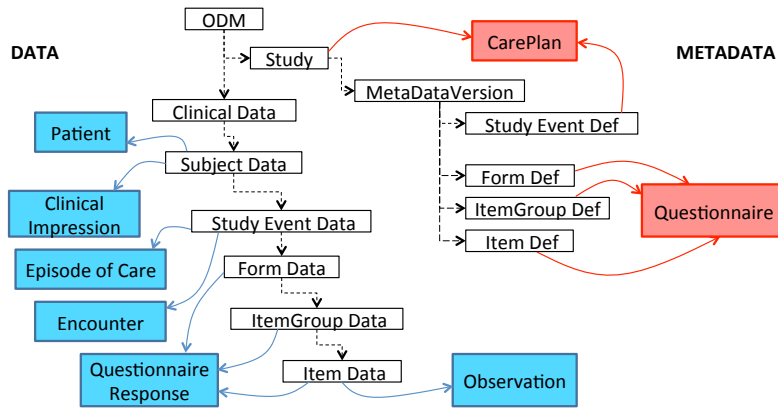


Fig. 3. Illustrates how the ODM entities are mapped to the FHIR resources

3.1 Study

A study defines static information about the structure of an individual study. We wish to not only capture this static information but also much of the contextual information pertaining to the study that is contained in the study protocol. This has resulted in the choice of the **CarePlan** resource to map the **Study** component from ODM. It provides a link to the study coordinator through the **participant** attribute and study protocol through the **support** attribute.

3.2 Subject

While the Subject represents a critical element of the study, its role is quite subdued in ODM. In particular, the specification provides no functionality to record the subject's attributes such as gender, date of birth, recommending that these be modelled as clinical data within the forms. The logical mapping for the Subject in FHIR is the **Patient** resource. This allows us to include relevant contextual information, such as the patient's gender, date of birth and care

provider and allows the study subject to be linked to other FHIR resources containing pertinent study-related information. The clinical data for each subject is encapsulated within a `ClinicalImpression` resource that is linked to the `Patient` resource.

3.3 Study Event

A study event comprises a `StudyEventDef` and a `StudyEventData` component that are referenced using a common OID. The `StudyEventDef` manages the set of forms to be completed at this phase of the study and represents an **activity** within the `CarePlan` resource. `StudyEventDef` entities define *scheduled* and *unscheduled* events and these may be defined within the `detail.scheduled` attribute of the activity. The `StudyEventData` entity contains clinical data collected during a subject's visit. We believe that the `EpisodeOfCare` resource is appropriate for this entity because it provides details about the group of activities and their purpose pertaining directly to a patient. The care plan is linked to this resource using the `plan` attribute. A study event may result in many visits from a patient. Each individual visit is modelled as an `Encounter` and is linked to the episode of care through the `episodeOfCare` attribute. The `patient` attribute links the resource to the study subject while the `assessor` attribute provides a link to the clinician conducting the clinical assessment.

3.4 Form

A form defines a collection of data items collected during the study and termed a *case report form*. A form comprises a `FormDef` and a `FormData` component that are referenced using a common OID. The form is linked to `CarePlan` through the `activity.actionResulting` attribute. The `FormDef` defines the form structure and its questions. The logical mapping of forms in FHIR is the `Questionnaire` (Q) resource. This resource contains the typical attributes for questionnaires, such as an identifier, version, publisher and status, but can also be customised using the extension mechanism in FHIR. The `FormData` entity contains the clinical data associated with the form. The logical mapping for the `FormData` in FHIR is the `QuestionnaireResponse` (QR) resource. The benefits of using the QR resource are that the order of the responses is maintained and these can be linked and validated against the questions asked. Conversely, however, while CDISC defines the CDASH³ model to standardise the generation of CRFs for clinical studies, its use is not enforced. Consequently, in our experience, few study coordinators choose to use them [11]. As a result, CRFs often contain contextually unrelated questions grouped together because they match the way in which the data entry person collects the data. Choosing to model this in a FHIR resource will only serve to perpetuate a bad practice [11].

³ Clinical Data Acquisition Standards Harmonization

3.5 Item Group

The `ItemGroupDef` and `ItemGroupData` entities constitute an item group referenced using a common OID. The `ItemGroupDef` entity defines the optional grouping of questions on a form. Groups are defined using the `Q.group` attribute. The FHIR specification stipulates that a group attribute define either a question or a group but not both. The `ItemGroupData` contains the clinical data detailing the responses for the item group. FHIR organises these grouped responses within the `QR.group` attribute. The FHIR specifications requires the order of the responses within the group to be maintained. This is a very important constraint. Consider the situation where the *heart rate* measurement of a patient indicates that the patient might have had a slight malaise during data collection and that subsequently a blood pressure measurement was taken. In this case, it would be prudent to analyse the blood pressure observation in this context and make inferences accordingly. On the flip side, questions are often grouped together on a form to match the collection habit of the data entry person and not necessarily because of their semantic similarity. Grouping responses in this manner does not advance the semantic interoperability principles.

3.6 Item

At the item level, the `ItemDef` and `ItemData` entities define each question and its subsequent response. The `ItemDef` entity defines the question asked during the study along with defining attributes such as the datatype, data size, measurement unit, permissible range and code list. The `Q.group.question` attribute is the most appropriate to define the `ItemDef` entity. The logical mapping for the `ItemData` entity is the `QR.group.question` attribute. The response to the question is then contained within the `question.answer` sub-attribute. This model works best in a lifestyle study scenario using questionnaires in the traditional *question-answer* mode. In the case of longitudinal clinical studies where the responses are analogous to a patient's observations during an episode of care, we believe the `ItemData` entity to be more appropriately represented using the `Observation` resource. Furthermore, as outlined in the FHIR specifications, data captured in questionnaires can be difficult to query after the fact. Individual items within a `QR` or an `Observation` are subsequently linked back to the `Encounter` in which they occur.

4 Discussion and Related Work

The FHIR resources provide a good fit to semantically enrich the extracted data from the CDISC ODM. In spite of its shortcomings in providing context to the clinical data, the CDISC ODM provides a sound hierarchical framework for capturing the clinical data that needs to be replicated in the new model. Several assumptions have been made when mapping the ODM data to the FHIR resources because their objectives differ and this work represents *one* view of

how the mapping can be achieved. We chose to model the study as a **CarePlan** because we want to model the activities planned for the patient during the study in the context of the study protocol. The **CarePlan** resource offers a number of attributes, such as **context**, **category** and **description** that can provide additional context to the care plan. The clinical data pertinent to a patient is modelled using the **ClinicalImpression** resource. The **ClinicalImpression** permits very pertinent information to be associated to the patient's data through the use of the **trigger**, **investigations** and **summary** attributes. Furthermore, it makes it possible to explicitly link the **protocol** followed and to associate the **findings** to the clinical data.

The **StudyEventDef** is modelled as an **activity** within the care plan. At a macro level, the study event data is categorised as an **EpisodeOfCare**. The **EpisodeOfCare** resource provides broad context to the study event. At a micro level, each visit within the study event is represented as an **Encounter**. This provides a richer summary of the activity performed, allowing each visit to be described atomically and linked back to the event using the **episodeOfCare** attribute. However, this requires the ODM data to be rich and accurate enough. This is particularly important for study event activities that occur at differing times in a particular day.

The **Questionnaire** resource is a suitable match for mapping ODM forms. However, as outlined in section 3.4, forms are often ill-conceived in ODM and as discussed in sections 3.5 and 3.6, the tendency is not to organise questions in a contextual manner but in one that befits the data capture process. The implication is that tremendous effort, which grows exponentially with the size of the study, must be expended to *semantically enrich* the clinical data by regrouping it contextually and integrating it with the relevant domain ontologies [10, 9].

As outlined in section 3.6, we advocate the use of the **Observation** resource to model the responses from item data. FHIR considers the **Questionnaire** resource to be a *specialisation* of the **Observation** resource. The appeal in adopting the **Observation** resource to store the **ItemData** responses is the ability to store important contextual information alongside the clinical data. Specifically, the **dataAbsentReason** attribute, which enables some justification to be provided as to the absence of a measurement, is very important in a clinical study setting, as is the ability to interpret the observation in the context of a controlled vocabulary or ontology. Furthermore, an **Observation** allows pertinent information such the **method**, **specimen** and **performer** but also the **device**, **bodySite** and **related** attributes to be coupled with the data, which is very useful when observing clinical data such as *vital signs*.

Ultimately, the aim of this work is to stimulate a debate on the most effective way to model clinical study data. While mapping ODM to FHIR is, in our view, a suitable solution, we do not regard it as a permanent one. Fundamentally, FHIR has the potential to manage clinical data in its own right. This has numerous advantages because as discussed in [7], a mapping process invariably leads to the loss of pertinent information. Furthermore, the process involves the reintroduction of critical domain information into the model. A more efficient

process would be to include this information *in casu* during the data collection phase.

4.1 Related Work

Several researchers have initiated approaches to address the semantic enrichment of clinical data with a view to achieving interoperability. One such approach, the Linked Clinical Data Cube (LCDC) [10, 9, 11] is a set of modularised data cubes that helps manage the multi-dimensional and multi-disciplinary nature of clinical data. A comprehensive comparison between the LCDC and this work is outside the scope of this paper. They have similar aims in trying to semantically enrich clinical study data and provide additional dimensions to overcome the monolithic nature of the ODM data and facilitate the exploration and querying of the data. The LCDC achieves this by introducing *specialised cubes*, *slices* and *observations*. In this work, all the chosen resources, except for the care plan and the questionnaires, have a link back to the patient. Furthermore, there are links between the Observation, Encounter, Questionnaire and QuestionnaireResponse resources to augment the dimensions offered. The *patient* and the *observation* are the main focus in both cases and they both provide the mechanisms to interpret the observations in the context of externally controlled vocabularies or ontologies. Furthermore, they both offer the ability to specify that the observation data is missing, although the **Observation** resource in FHIR also includes the justification as to why the result is missing. The main difference is in the approach. The LCDC requires mapping to the RDF Data Cube [5] and DDI-RDF Discovery [3] vocabularies to organise the data and links to domain ontologies to enrich it. In this work, the organisation, management and enrichment are performed by FHIR using **Codeable** concepts to link it externally.

Dugas [7] describes two tools to convert forms between the CDISC ODM and HL7 CDA⁴ formats to facilitate the sharing of electronic health records (EHRs) and clinical data to address the problem of redundant documentation in both systems. He concluded that the conversion process is *lossy* because the CDISC and HL7 models serve different purposes and hence have different properties. Similarly, the SALUS project [6] aims to address the interoperability between clinical care and the clinical research domain. More specifically, it looks at combining the strengths of CRFs with those of EHRs to address adverse drug reactions.

Abler et al. [1] discuss the need for a language for forms that can effectively record the logical relationships between questions or sets of questions asked in the forms. While the natural inclination would be to look to the **Questionnaire** resource to fulfil this need, a more encompassing solution would be to integrate the capabilities of the **Observation** resource with the questionnaire.

The Pharmaceutical Users Software Exchange⁵ community, in concert with the FDA, has started work on RDF representations of various CDISC models⁶,

⁴ Clinical Document Architecture

⁵ <http://www.phusewiki.org/wiki/index.php?title=Semantic.Technology>

⁶ <https://github.com/phuse-org/rdf.cdisc.org>

including the terminologies published by the National Cancer Institute (NCI) Enterprise Vocabulary Services⁷. This community has started to evaluate the RDF Data Cube [2, 13] for the publication of clinical study data.

The HL7 Working group on Semantic Interoperability⁸ has initiated some work on translating the XML or JSON version of FHIR into FHIR RDF. This work is still in *draft* mode and we will look at translating our solution into FHIR RDF once it has reached a more mature level.

5 Conclusion

As secondary use of clinical data gathers momentum, it will become increasingly important to share and compare clinical studies. We have presented an approach that integrates clinical data extracted from CDISC ODM to the FHIR resources. This work takes advantage of the rich information model comprising the FHIR resources to *semantically enrich* the clinical data and reintroduce the domain information that was omitted during the data capture phase. The objective is to achieve semantic interoperability of clinical study data by standardising and normalising the data along the same metrics. The main contribution is a framework to organise clinical data in a manner that preserves its organisation but captures its context. A sample of this mapping is available at: <http://healthinet.it.csiro.au/net/odmFhirMapping/>.

References

1. Abler, D., Crichton, C., Welch, J., Davies, J., Harris, S.: Models for forms. In: Proceedings of the compilation of the co-located workshops on DSM'11, TMC'11, AGERE'11, AOOPEs'11, NEAT'11, & VMIL'11. pp. 13–18. ACM (2011)
2. Andersen, M.: Linked data to support clinical and non-clinical reporting. In: Proceedings of 2nd International Workshop on Semantic Statistics (SemStats 2014). Riva del Garda, Italy (2014)
3. Bosch, T., Cyganiak, R., Gregory, A., Wackerow, J.: DDI-RDF Discovery Vocabulary: A metadata vocabulary for documenting research and survey data. In: LDOW (2013)
4. CDISC: Specification for the Operational Data Model (ODM) (2006), <http://www.cdisc.org/models/odm/v1.3/ODM1-3-0-Final.html>
5. Cyganiak, R., Reynolds, D., Tennison, J.: The RDF Data Cube Vocabulary (2013), <http://www.w3.org/TR/2013/PR-vocab-data-cube-20131217/>
6. Declerck, G., Hussain, S., Daniel, C., Yuksel, M., Laleci, G.B., Twagirumukiza, M., Jaulent, M.C.: Bridging data models and terminologies to support adverse drug event reporting using EHR data. *Methods of Information in Medicine* 54(1), 24–31 (2015)
7. Dugas, M.: ODM2CDA and CDA2ODM: Tools to convert documentation forms between EDC and EHR systems. *BMC medical informatics and decision making* 15(1), 40 (2015)

⁷ <http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources/cdisc>

⁸ http://wiki.hl7.org/index.php?title=RDF_for_Semantic_Interoperability

8. Hsu, W., Gonzalez, N.R., Chien, A., Villablanca, J.P., Pajukanta, P., Viñuela, F., Bui, A.A.: An integrated, ontology-driven approach to constructing observational databases for research. *Journal of biomedical informatics* 55, 132–142 (2015)
9. Lefort, L., Leroux, H.: Design and generation of linked clinical data cubes. In: *Proceedings of 1st International Workshop on Semantic Statistics (SemStats 2013)*. Sydney, Australia (2013)
10. Leroux, H., Lefort, L.: Using CDISC ODM and the RDF Data Cube for the semantic enrichment of longitudinal clinical trial data. In: Paschke, A., Burger, A., Romano, P., Marshall, M.S., Splendiani, A. (eds.) *Semantic Web Applications and Tools for the Life Sciences (SWAT4LS)*. CEUR Proceedings (2012)
11. Leroux, H., Lefort, L.: Semantic enrichment of longitudinal clinical study data using the CDISC standards and the semantic statistics vocabularies. *Journal of Biomedical Semantics* 6(1), 16 (2015)
12. van Valkenhoef, G., Tervonen, T., de Brock, B., Hillege, H.: Deficiencies in the transfer and availability of clinical trials evidence: a review of existing systems and standards. *BMC medical informatics and decision making* 12(1), 95 (2012)
13. Williams, T.: A primer on converting analysis results data to RDF Data Cubes using free and open source tools. In: *Proceedings of 10th Annual PhUSE conference(PhUSE 2014)*. London, UK (2014)

A Appendix

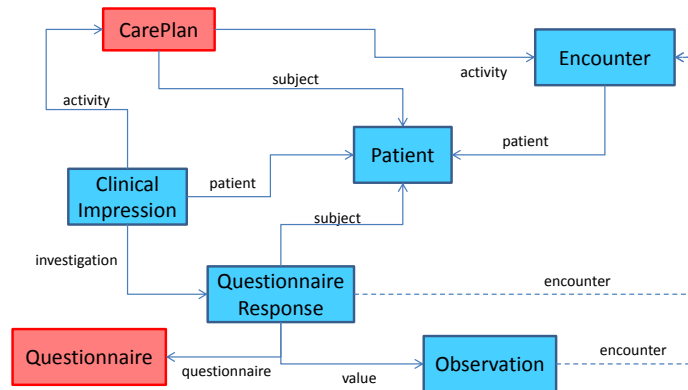


Fig. 4. Original model depicting the *metadata* (red) and *data* (blue) FHIR resources with actual (solid) and potential (dashed) links