Using CDISC ODM and the RDF Data Cube for the Semantic Enrichment of Longitudinal Clinical Trial Data

Hugo Leroux[1] and Laurent Lefort[2]

[1]The Australian E-Health Research Centre, CSIRO, Brisbane, Queensland, Australia
[2]CSIRO ICT Centre, Canberra, ACT, Australia
{firstname.lastname}@csiro.au

Abstract. The development of ontology, linked data standards and tools for semantic enrichment, opens new opportunities to analyse and reuse the clinical data collected as part of clinical trials and longitudinal studies. This paper presents our approach on the semantic enrichment of the data collected as part of the Australian, Imaging, Biomarker and Lifestyle study of Ageing (AIBL). AIBL is a large scale longitudinal clinical study into neurodegenerative diseases that has been designed to support investigations of the predictive utility of various biomarkers, cognitive parameters and lifestyle factors as indicators of Alzheimer's disease.

The objective of this paper is to highlight the complementarities of Clinical Data Management Systems standards, such as CDISC ODM, with novel approaches to manage large volume of heterogeneous linked data resources, such as the W3C RDF Data Cube. We start by describing the standards, ontologies, linked data resources and tools that we use to aggregate the study data. Next, we describe the structure of the Linked Clinical Data Cube and the tools that we use to map the recorded medication intake information to the relevant standards in the Australian context: SNOMED CT and the Australian Medical Terminology. We also discuss how our approach could be extended to take advantage of past and present Linked Open Data initiatives in the Health Care and Life Sciences community.

Keywords: Ontology, Semantic enrichment, Clinical Trial, Longitudinal Study, Medication data, Data Cube

1 Introduction

The implementation of clinical trial management systems using relational databases has expedited the dissemination of clinical trial data among collaborators and provided the potential to swiftly query the data repository to extract information. Existing Clinical Data Management Systems (CDMS) software use a generic electronic case report form (CRF) data structure to adapt to multiple clinical trials, and are implemented in relational database management systems through a set of monolithic tables.

SWAT4LS-2012, November 28-30, Paris, France.

Some CDMS products such as OpenClinica[1] use a data dictionary, derived from the clinical study protocol, to handle the variable fields of these generic tables. These products generally support the Clinical Data Interchange Standards Consortium Operational Data Model (CDISC ODM) export file format [1], an XML-based standard that specifies how to include this lightweight study metamodel as metadata alongside the study data.

The RDF Data Cube vocabulary [2], developed by the W3C Government Linked Data working group, is a vocabulary for the publication of statistical data in RDF. The specification defines a generic *observation* data structure that matches the CRF data structure used in clinical data management systems.

In this work, we present how we can apply the RDF Data Cube specification to semantically enrich longitudinal clinical study data to allow users to query the clinical trial data more effectively and efficiently. We describe how the proposed *Linked Clinical Data Cube* can support the various case report forms derived from the study protocol in a uniform manner.

The approach proposed in [3] is evaluated on clinical data obtained as part of the Australian, Imaging, Biomarker and Lifestyle study of Ageing (AIBL) data. AIBL [4] is a large scale longitudinal clinical study into neurodegenerative diseases that has been designed to support investigations of the predictive utility of various biomarkers, cognitive parameters and lifestyle factors as indicators of Alzheimer's disease (AD) with a cohort of over one thousand participants residing in two Australian cities, Perth and Melbourne. Each recruited participant completed blood and neurological testing and some underwent brain imaging testing. To fulfil the longitudinal nature of the study, each participant undergoes a re-examination every eighteen months. The participants also volunteered a broad range of lifestyle and medical information, including medication information that is currently obtained from participants in paper form and then manually entered into the OpenClinica CDMS by study staff [5].

For the enrichment of medication data, we will use the Australian Medicines Terminology (AMT) and SNOMED CT-AU [6]. However, among the difficulties associated with collating self-reported medication usage is the potential for misidentification of the correct medication being recorded along with inconsistencies and lack of precision relating to crucial information, such as dosage and frequency of use. We show here how we can extend our Linked Clinical Data Cube to reuse links between medication resources [7] and in particular reuse the ones which are already available as linked data [8-9] or which share common identification keys [10].

This paper has four main sections: in section 2, we introduce and compare the CDISC ODM and of the RDF Data Cube specifications. In section 3, we describe the generic structure of the Linked Clinical Data Cube. In section 4, we describe in more details how we can convert the recorded medication intake information recorded for the AIBL study and map it to the relevant standards in the Australian context: SNOMED CT and the Australian Medical Terminology. Finally, in section 5, we discuss how our approach could be extended to take advantage of past and present Linked Open Data initiatives in the Health Care and Life Sciences community.

[1] https://openclinica.com/

2 Comparison of CDISC ODM and RDF Data Cube

2.1 CDISC ODM (with OpenClinica extensions)

The AIBL study data was migrated from a proprietary CDMS to the OpenClinica CDMS a year ago [5]. One of the compelling factors in choosing OpenClinica was that it adopts the CDISC ODM structure to define the logical organisation of the study metamodel and as the basis for the standard export format for the clinical trial data. The CDISC ODM standard [1] defines a format that facilitates the sharing of clinical data and metadata from multiple sources. Furthermore, it assumes that the clinical study follows a predefined structure that defines the study, subject, events, forms, item groups and items. An abstracted version of the CDISC ODM structure that illustrates the regions of interest for our project is shown in **Fig. 1**.

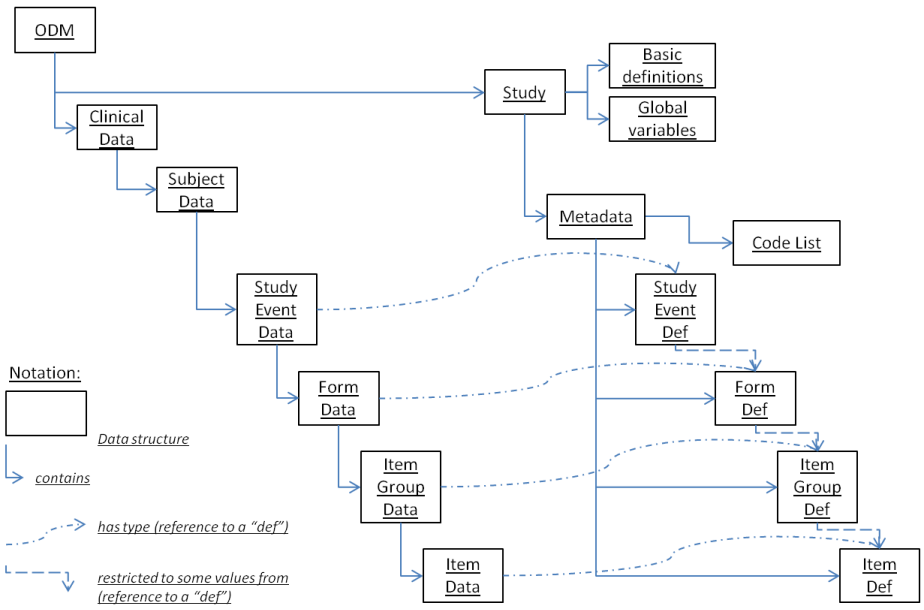

Fig. 1. The OpenClinica CDISC ODM data model (container and descriptor classes)

The CDISC ODM data model is specifically designed for a data capture context:

- An item is a single measurement or analysis result collected during a study e.g. a blood pressure reading,
- An item group is a set of related measurements or analysis results,
- A form (or case report form) is a collection of items and item groups for capturing and displaying clinical trial data,
- A study event corresponds to a patient visit or other encounter where the data corresponding to one or multiple forms is collected.

Furthermore, the definitions of these data structures are also included in the CDISC ODM export format. In OpenClinica, this data dictionary can be edited by super users with the help of an Excel template that plays the role of a configuration file. The users of the tools are encouraged to share their CRFs and have access to a library of peer reviewed ones derived from authoritative standards sources such as the CDISC Clinical Data Acquisition Standards Harmonization (CDASH) initiative.

2.2 The RDF Data Cube vocabulary

The RDF Data Cube vocabulary [2], published by the W3C Government Linked Data working group, is a vocabulary for the publication of statistical data in RDF [11]. This specification is available as a working draft, but it has been evaluated by a number of government agencies (Eurostat, European and UK Environment agencies) who have published large scale datasets. It has also triggered new work on the Online Analytical Processing (OLAP) of Linked Data sources ([12-13]).
The basic principles behind the design of the RDF Data Cube vocabulary are illustrated in Figure 2.

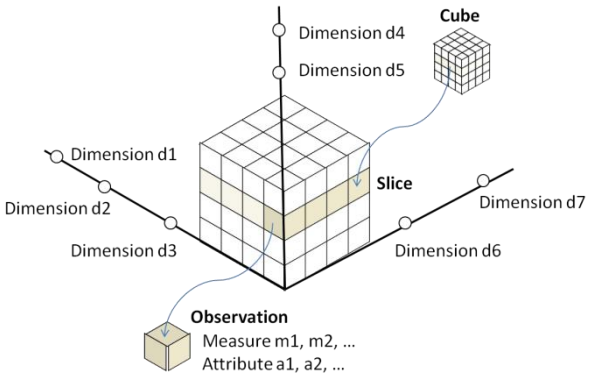

Fig. 2. Cube (Dataset), Slice and Observation

A cube is a *dataset* that is divided into *slices* according to several dimensions. Each slice contains a number of *observations*. The arrows in **Fig. 2** represent the links between the cube and the slices and between the slices and the observations. These extra links at multiple levels of data aggregation allow the data consumers to navigate and query linked data. The RDF Data Cube vocabulary defines three types of data items: *dimensions* for the identification keys, *measures* and *attributes* for the recorded data and metadata. The slices group subsets of observations within a dataset where all the dimensions except one (or a small number) are fixed.
The RDF Data Cube vocabulary specifies container classes and descriptor classes and the set of properties to link them (**Fig. 3**). The main descriptor class, has a link to properties classes (`qb:componentProperty`) to specify the data items which are used. The other classes (`qb:DataStructureDefinition`, `qb:SliceKey`, `qb:ComponentSpecification`) and properties (`qb:structure`, `qb:sliceStructure`, `qb:componentAttachment`) are used to indicate what

properties are used at what level of aggregation (`qb:Observation`, `qb:Slice` and `qb:DataSet`).

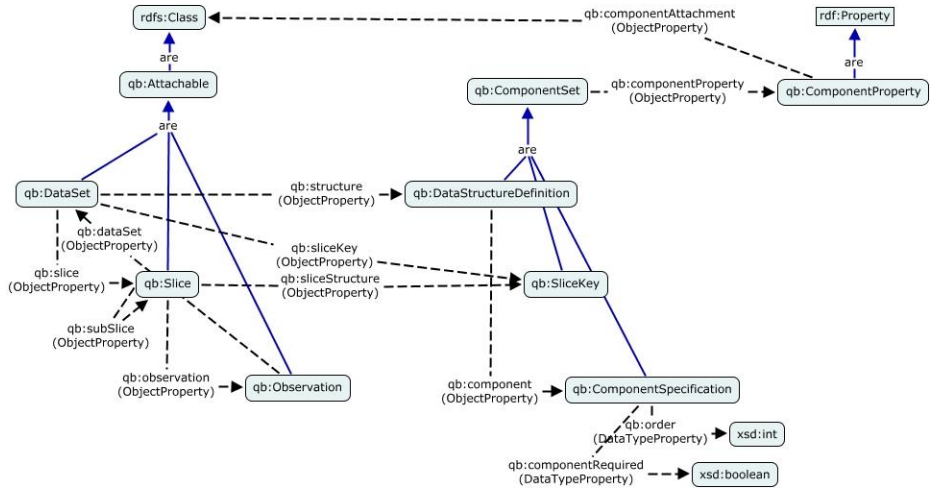

Fig. 3. The RDF Data Cube data model (container and descriptor classes)

2.3 Comparison of data slicing approaches

Both approaches bundle together the data (container classes) and metadata (descriptor classes) and support the registration of code lists. The CDISC ODM format supports only one way of slicing the data with five primary dimensions (Subject-Study event-Form-Item group-Item). The RDF Data Cube structure is more flexible for later publication of the collected data where the choice and ordering of the dimensions depend largely on the queries that we wish to make from the system.

In the context of the AIBL study, as it is longitudinal in nature, we are particularly interested in analysing the different variables from an *Event* perspective; for example to determine the rate of change of a particular observation across different time points. On the other hand, we are also interested in analysing variables at an *Item* level. One such example would be to track the concentration of a particular protein in the blood together with the cortical thickness of a portion of the brain.

3 Structure of the Linked Clinical Data Cube

3.1 Overview

We use a Linked Clinical Data Cube template that answers our needs for analysing the AIBL study from an Event and an Item perspective. The primary dimensions for the CDISC ODM data model (**Fig. 1**) are the subject (or patient) identification and the study event (or time point) identification. The other dimensions (Form – Item group

and Item) are domain-dependent and are specified by the data dictionary content. To provide access to the recorded data at both levels, we have designed a Linked Clinical Data Cube based on two nested RDF Data Cubes as shown in **Fig. 4**.

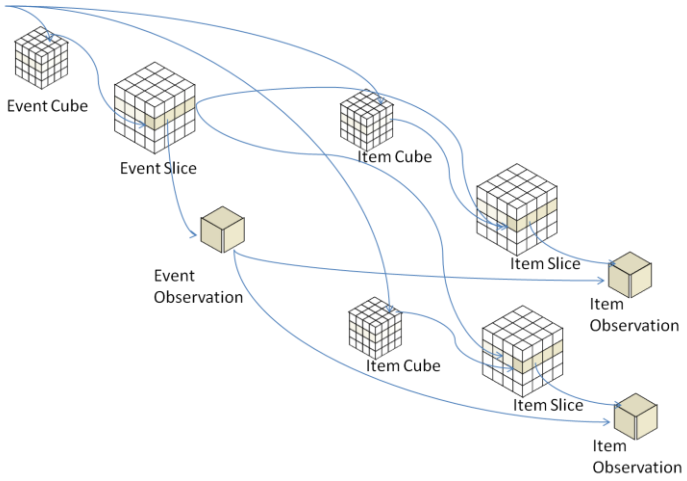

Fig. 4. Event and Item data cubes

This Linked Clinical Data Cube comprises two nested data cubes that depict the interconnectedness between an *Event Cube*, an *Event Slice*, an *Item Cube* and an *Item Slice*. Our top-level data cube manages slices of `Study Event Observations` that contain observations that are collection of Form (CRF) data. Our bottom-level data cube manages slices of `Item Observations` that contain observations that are either Form data or item data. To minimize the duplication of data, the Event Observations contains links to the low-level Item Observations containing the measures and attributes rather than their values. This provides us with a high level of flexibility to analyse observations at both an Event and Item level concurrently.

3.2 Coupling with domain ontologies

The RDF Data Cube vocabulary (QB) has been coupled to domain ontologies for the publication of long term climate data time series as linked data [14] with the help of the W3C Semantic Sensor Network ontology[2] [15]. For the Linked Clinical Data Cube, we will need different ontologies for each item data cube corresponding to a different domain of application.

Users of CDISC-compliant tools are encouraged to use standard Case Report Forms (CRFs) to directly comply with other CDISC standards such as the CDISC Clinical Data Acquisition Standards Harmonization [16] and the CDISC Study Data Tabulation Model [17] specifications. SDTM and CDASH information can also be added to

[2] `http://purl.oclc.org/NET/ssnx/ssn`

CDISC ODM content as annotations (using the Alias element) at all levels of definitions (for all the descriptor classes shown in Figure 1).
Fig. 5 illustrates how we can potentially integrate a lightweight "skeleton ontology" based on these CDISC standards with ontology modules from the RDF Data Cube vocabulary (classes with the `qb` prefix) and SSN ontology (classes with the `ssn` prefix) to construct our Linked Clinical Data Cube.

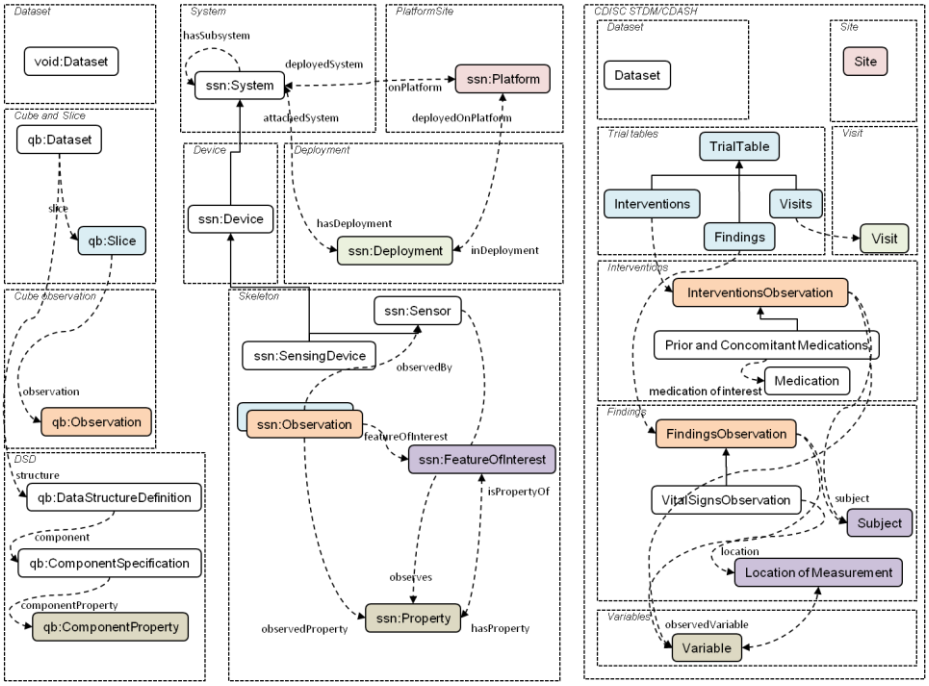

Fig. 5. Coupling the AIBL Linked Clinical Data Cube with CDISC-mappable ontologies

Fig. 5 shows the ontology modules, their classes and relationships (plain lines are used for sub-class-of relationships and dashed lines for object properties linking classes). The different colours used in **Fig. 5** indicate which classes from which ontology modules should be coupled together (e.g. via multiple inheritance relationships):

- `qb:Slice`, `ssn:Observation` and `TrialTable` (`Interventions`, `Findings` and `Visits`),
- `qb:Observation`, `ssn:Observation`, `InterventionsObservations` and `FindingObservations`,
- `qb:ComponentProperty`, `ssn:Property` and `Variable`,
- `ssn:FeatureOfInterest`, `Subject` and `LocationOfMeasurement`,
- `ssn:Platform` and `Site`,
- `ssn:Deployment` and `Visit`.

4 Application to the AIBL medication data

4.1 Data collected for the AIBL study

The AIBL study has been collecting medications information in part to monitor the effects of some pharmaceuticals that could affect cognitive function. Information relating to the medications intake of each participant was recorded on a questionnaire, in paper form, and manually entered, by study staff, in OpenClinica. A sample of the recorded information including the medication's *name*, prescribed *dosage*, *frequency* and *duration* of use is shown in **Table 1**.

Table 1. The medication information as recorded in OpenClinica

Subject id	Study Event id	Item group id	Medication name	Dosage	Frequency	Length of time taken
4	3	3	Cartia			3 years
26	3	1	Arthro-aid (glucosamine hydrochloride)	750mg	1 bd	

The goal is to map this medication information to taxonomy of medication codes in order to provide a hierarchical classification of the drugs. One significant challenge linked to this task is in the identification of the correct medication given the inaccuracy due to inconsistency with the naming and imprecision regarding the dosage, frequency and duration of use. In the case of the medication's name, a mix of *Trade Name*, *Active Ingredients* and *informal name* have been used to describe the prescribed medication. Furthermore, the participants have omitted to record several fields including the prescribed dosage, frequency and duration of use when filling in their questionnaires. In the next section, we describe our approach to mapping the medication information to two Australian standards for medication terminology: AMT and SNOMED CT-AU.

4.2 Mapping to SNOMED and AMT

Our choice of AMT and SNOMED CT is based on their complementarities. AMT provides unique codes and accurate standardised names to unambiguously identify all commonly used medicines in Australia with eight key top-level concepts [18] including Trade Product. SNOMED CT organises content into several hierarchies, including the Substance, Clinical finding, Body structure and Observable entity hierarchy and its foundation in Description Logic makes it a good candidate to decomposing the complex medications concept hierarchy and describing our domain ontology.
The processing pipeline [6] for mapping the medication information is shown in **Fig. 6** and summarised below.

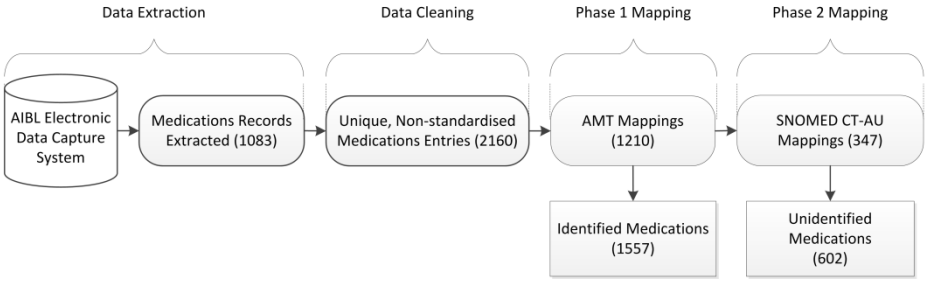

Fig. 6. Processing pipeline for mapping the medications data (extract from [6])

The medication records are extracted from OpenClinica at the start of the pipeline. A data cleansing process is conducted to manually amend the inconsistencies, described in the previous section, from these records. This is followed by two mapping phases. In Phase 1, the system attempts a match of the medication name to an AMT concept below the Trade Product hierarchy. The search operation returns zero or more candidate mappings. If more than one concept is returned, the strategy adopted to match the AIBL medication to an AMT concept is to calculate the Least Common Ancestor (LCA) [6]. During Phase 2, for every medication name not adequately identified in Phase 1, the system attempts a match to a SNOMED CT-AU Substance Identifier. The use of the Substance hierarchy is designed to broaden the search in an attempt to address the more obscure medication name not identified in Phase 1 [6].

4.3 Handling AIBL medication records in the Linked Clinical Data Cube

The Medication Data Cube is an instance of the Item data cube described in **Fig. 4**. Its primary dimensions are the subject id and study event id. The originally available dimension for the Medication reference is the Medication name. The AMT and SNOMED-CT identifiers can be used as alternative dimensions when available as described in **Fig. 7**. The name, dosage, frequency and duration of use are available as measures or attributes.

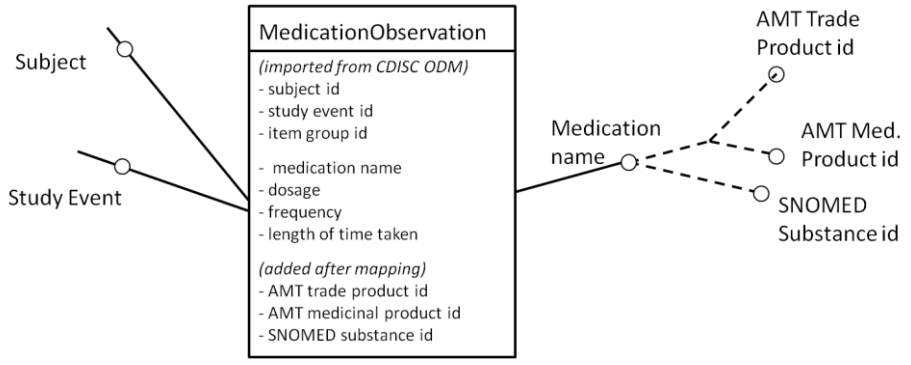

Fig. 7. Data cube dimensions for the AIBL Medication Data

Fig. 8 extends the discussion from section 3.2 by illustrating how the references from the SNOMED CT and AMT ontologies augment the skeleton ontology depicted in **Fig. 5**. Linking to AMT and SNOMED CT concepts provide the possibility to obtain additional information based on links between the concepts or trade products branches and other branches in the AMT and SNOMED CT ontologies. We will also exploit the mappings at the substance level between these two resources as defined in [18].

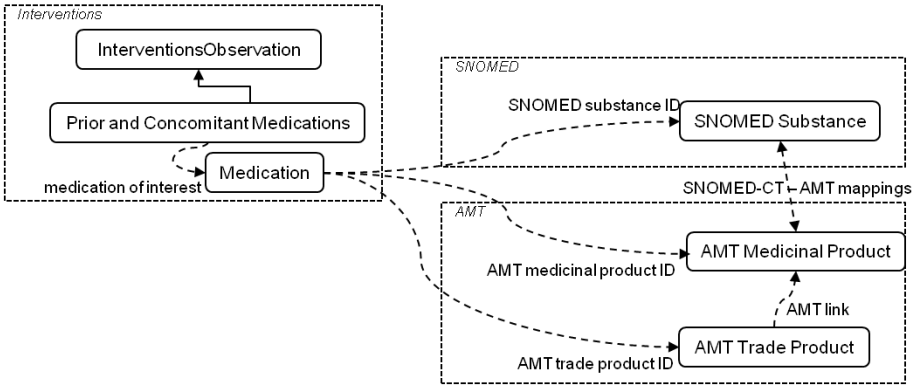

Fig. 8. Medication observation reference to SNOMED or AMT ontologies

We also intend to use the DrugBank[3,4] database and the Anatomical Therapeutic Chemical (ATC) and Defined Daily Dose (DDD) taxonomy[5] defined by the WHO Collaborating Centre for Drug Statistics Methodology to supplement the medication data as depicted below in **Fig. 9**.

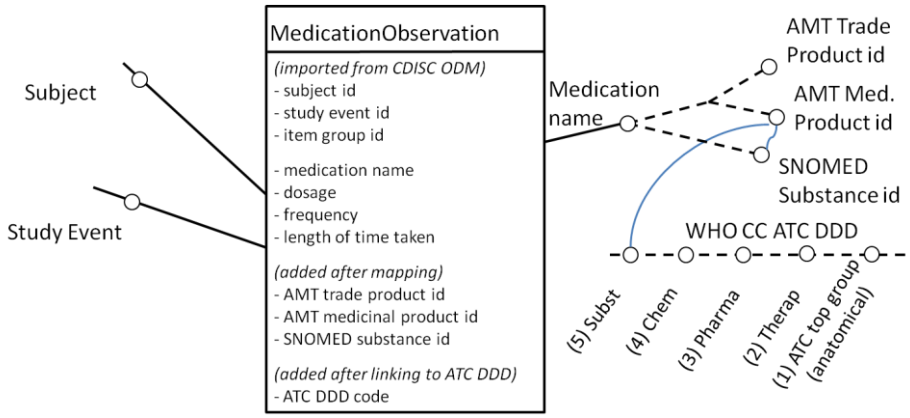

Fig. 9. Additional dimensions resulting from the linking to ATC DDD

[3] http://www.drugbank.ca/
[4] A RDF version of DrugBank is available from http://linkedlifedata.com/.
[5] http://www.whocc.no/atc_ddd_index/

5 Discussion

5.1 Benefits of the approach

There are several challenges associated with mapping clinical trial concepts to established ontologies and linked data resources to enrich clinical data such as the AIBL study data [3]. We propose a three-tiered approach which helps to answer some of these challenges.

The first tier applies the Data Cube principles to overcome the monolithic nature of the CDISC ODM file structure. This is the approach illustrated by **Fig. 4** which exposes the clinical data across multiple dimensions.

The second tier involves the semantic enrichment of the AIBL data using references from the curated medication classification obtained by mapping the medication data to AMT and SNOMED CT. This is outlined in section 4.3 and illustrated by **Fig. 7**. This process has the potential to further expose the clinical data across the additional dimensions.

The third tier relates to the linkage of the clinical data to other resources, namely the ATC DDD, DrugBank and all the other linked data resources that possess references to them. This approach is depicted in **Fig. 9** and provides the opportunity to introduce yet supplementary dimensions through which to expose the data. For the users of the AIBL data published as linked data, the benefits of our approach are tied to the extra information provided by the linked resources as adding links to DrugBank and ATC DDD create new opportunities to query the data.

DrugBank also defines drug and food interactions. The former provides an important step in the exploration of *drug-drug* interactions that also provide some insight into potential risks and contraindications associated with the intake of the medication. The latter will be useful when we explore the association between the participant's drug intake and type and quantity of food consumed. DrugBank also provides information on the *gene-drug* interactions medication target which could expedite the discovery of biomarkers.

The five levels of taxonomy of medications code provided by ATC DDD (**Fig. 9**) also provide means to aggregate the study data for statistical purposes. This is complementary to what would be possible with the help of the taxonomies supplied by AMT and SNOMED-CT.

5.2 Future work

The Linked Clinical Data Cube will not reach its true potential unless it is coupled with multiple domain ontologies to enrich its referencing capabilities. The work within the AIBL Linked Clinical Data Cube will be to organise and logically map the logical information contained within the various CRFs to domain ontologies (Fig. 1 in [5]). We plan, in the short term, to conduct a survey of existing domain ontologies, from the literature, to identify suitable candidates that adequately define the semantics of the test data comprising the study.

As a first step, we will need to identify the primary dimensions and the set of identifiable classes that define the Linked Clinical Data Cube. For the first tier, we need a modular ontology that covers the definitions introduced by CDISC standards, in particular CDASH [16]. As shown in **Fig. 5**, the skeleton of this ontology can reuse a good subset of the Semantic Sensor Network ontology but it should also define key CDASH classes such as `Intervention`, `Findings`, `Visit` and `Subject`. We also need additional modules for each type of CRF defined for the AIBL study data [5]. One of the roadblocks to this task is the need to release RDF versions of the CDISC CDASH [16] and STDM [17] standards in sync with versions of these standards used in the tools. To ease the conversion from CDISC ODM to RDF and encourage developers of new CRFs to map their definitions to a common reference, the reusable CRF templates supplied by the CDISC consortium should also include annotations pointing to CDASH definitions published as RDF. There have been several attempts by the CDISC Consortium to develop an RDF version of these two standards but these have, as yet, not been completed.

The second step entails producing more complete mapping tables between our concepts and those defined in linkable resources on the web, in particular AMT and SNOMED CT. There are opportunities to improve the semi-automated mapping algorithm implemented for AMT and SNOMED CT with the help of other medication resources e.g. DrugBank, NDF-RT[6] and RxNorm[7]. Schulz [19] identifies various shortcomings within SNOMED CT in relation to completeness and consistency.

5.3 Related work

Many researchers have developed approaches to facilitate the semantic enrichment of biomedical research data. Some of these approaches [20] have focussed on integrating the clinical data with ontologies while other approaches [21] have investigated the use of linked data resources. However, little effort has been directed at combining these two complementary approaches.

Some of the ontologies developed in the context of translational research [22] and clinical trials [23-25] are partially applicable to our needs. But they do not adequately cover the observation aspects that are required for our data cube. Several of these ontologies also have a large number of dependencies to other ontologies.

The Linked Open Drug Data[8] (LODD) and the Linked Life Data (LLD) projects provide additional resources that can be used to extend the AIBL Linked Clinical Data Cube. Both projects aim to build a large scale knowledge cloud that can be used for drug discovery. LODD [8] federates the efforts by participants of the W3C Health and Life Sciences (HCLS) Interest group[9] to convert available resources into linked data. LLD [9] provides a semantic data integration platform for the biomedical domain comprising many of the data sources belonging to LODD plus some new ones. The

---

[6] `http://evs.nci.nih.gov/ftp1/NDF-RT/`

[7] `https://www.nlm.nih.gov/research/umls/rxnorm/`

[8] `http://www.w3.org/wiki/HCLSIG/LODD/Data`

[9] `http://www.w3.org/wiki/HCLSIG`

resulting datasets contains more than 8 million triples representing the knowledge within over 2 millions links relating to medications, diseases, clinical trials, gene information and pharmaceutical companies among others.
Among the various use cases reported via the W3C HCLS Interest group are efforts to explore links to identify and verify genes linked to Alzheimer's disease (AD). Through the links between the drug, medications, disease and clinical trial repositories, we hope to leverage on efforts by others to further explore the effects of prescribed medications, for AD sufferers, on the various genes comprising the pathways of interest. Other applications of LODD include the identification of potential side-effects linked to the intake of drugs that have conflicting stimuli on the disease pathways.
The SALUS project [26] is a former attempt to adapt CDISC standards to build a Semantic Framework to improve interoperability between clinical research and clinical care domains. We adopt a similar approach to them but their focus is on service mappings rather than linked data sets. The Semantic Cockpit [27] project aims to develop a data slicing framework comparable to what we propose on the basis of the RDF Data Cube. The goal of this project is to intelligently assist business analysts by discriminating unimportant information and using reasoning to only present useful information to the analyst.

6 Conclusions

Several new opportunities exist to analyse and reuse the clinical data gathered as part of clinical trials through the development of ontology, linked data standards and tools to semantically enrich this data. We have presented an approach for the semantic enrichment of clinical trial data obtained as part of the AIBL study, a large-scale longitudinal study into neurodegenerative diseases. We have outlined the design of the *Linked Clinical Data Cube.* The *Linked Clinical Data Cube* takes advantage of the strength of the RDF Data Cube in defining the slices, dimensions and observations within the data and applying them to the CDISC ODM data model to provide increased flexibility in the formulation of queries and allow the users to query the clinical data more effectively and efficiently. We have also outlined the use of the AMT and SNOMED CT-AU taxonomies to enrich the medication data. Finally, we have presented our method to extend our Linked Clinical Data Cube to reuse links between medication resources, in particular the ones that are already available as linked open data. The main contribution of our approach is that we propose the use of ontologies and linked data resources together to semantically enrich the clinical data, thanks to the cohabitation of the *container* and *description* classes in our solution. Our strength is in the potential for semantic enrichment of any CDMS tools that adopts the CDISC standard.

Acknowledgements. The authors would like to express their gratitude to Simon McBride, Simon Gibson and Dr Michael Lawley for their assistance in scoping the

Medications case study and to Drs Alejandro Metke and Kerry Taylor for their feedback on the paper.

References

1. CDISC: Clinical Data Interchange Standards Consortium - Operational Data Model (2011) http://www.cdisc.org/odm
2. Cyganiak, R., Reynolds, D., Tennison, J.: The RDF Data Cube Vocabulary, W3C Working Draft 05 April 2012. World Wide Web Consortium (2012)
3. Leroux, H., McBride, S., Lefort, L., Kemp, M., Gibson, S.: A method for the semantic enrichment of clinical trial data. Stud Health Technol Inform, 178, pp. 111-116 (2012)
4. Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N. T., Lenzo, N., Martins, R. N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M. and Ames, D.: The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int. Psychogeriatry, 21(4), pp. 672-687 (2009)
5. Leroux, H., McBride, S., Gibson, S.: On selecting a clinical trial management system for large scale, multi-centre, multi-modal clinical research study. Studies in health technology and informatics, 168, pp. 89-95 (2011)
6. McBride, S., Lawley, M., Leroux, H., Gibson, S.: Using Australian Medicines Terminology (AMT) and SNOMED CT-AU to better support clinical research, Studies in Health Technology and Informatics, 178, pp. 144-149 (2012)
7. Sharp, M., Bodenreider, O., Wacholder, N.: A Framework for Characterizing Drug Information Sources. AMIA 2008 Symposium Proceedings (2008)
8. Samwald, M., Jentzsch, A., Bouton, C., Stie Kallesøe, C., Willighagen, E., Hajagos, J., Marshall, M. S., Prud'hommeaux, E., Hassenzadeh, O., Pichler, E.: Linked open drug data for pharmaceutical research and development. Journal of Cheminformatics, 3(1), pp. 19- (2011)
9. Williams, A. J.; Harland, L.; Groth, P.; Pettifer, S.; Chichester, C.; Willighagen, E. L.; Evelo, C. T.; Blomberg, N.; Ecker, G.; Goble, C. & Mons, B.: Open PHACTS: semantic interoperability for drug discovery. Drug Discovery Today (2012)
10. Saitwal, H. and Qing, D. and Jones, S. and Bernstam, E. and Chute, C.G. and Johnson, T.R.: Cross-terminology mapping challenges: A demonstration using medication. Journal of Biomed. Inform. 45, pp. 613-625 (2012)
11. Cyganiak, R., Hausenblas, M., McCuirc, E.: Official Statistics and the Practice of Data Fidelity. In: Wood, D. ed.: Linking Government Data. Springer, pp. 135-151 (2011)
12. Kampgen, B., O'Riain, S., Harth, A.: Interacting with Statistical Linked Data via OLAP Operations. In: International Workshop on Linked APIs for the Semantic Web (LAPIS 2012), (2012) http://lapis2012.linkedservices.org/
13. Etcheverry, L., Vaisman, A.A.: QB4OLAP: A Vocabulary for OLAP Cubes on the Semantic Web. In: Third International Workshop on Consuming Linked Data (COLD 2012), CEUR Workshop proceeding, vol. 905, CEUR-WS.org (2012),
14. Lefort, L., Bobruk, J., Haller, A., Taylor, K. and Woolf, A.: A Linked Sensor Data Cube for a 100 year homogenized daily temperature dataset. In: 5th International Workshop on Semantic Sensor Networks (SSN-2012), CEUR-Proceedings, vol. 904, CEUR-WS.org (2012).

15. Compton, M., Barnaghi, P., Bermudez, L., Garcia-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Phuoc, D. L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., Taylor, K.: The SSN ontology of the W3C semantic sensor network incubator group. Web Semantics: Science, Services and Agents on the World Wide Web 15(3), (2012)
16. CDISC CDASH Team: Clinical Data Acquisition Standards Harmonization (CDASH), Version 1.1 Clinical Data Interchange Standards Consortium (2011) http://www.cdisc.org/cdash
17. CDISC SDS Team: CDISC Study Data Tabulation Model, Version 1.2 Clinical Data Interchange Standards Consortium (2008) http://www.cdisc.org/stdm
18. Michel, J., Lawley, M. J., Chu, A., Barned, J.: Mapping the Queensland Health iPharmacy Medication File to the Australian Medicines Terminology Using Snapper. Studies in Health Technology and Informatics 168, pp. 104-116 (2011)
19. Schulz, S., Suntisrivaraporn, B., Baader, F., Boeker, M.: SNOMED reaching its adolescence: Ontologists' and logicians' health check. Int. J. Med. Inform., 78, pp. S86-S94 (2008)
20. Dumontier, M., Villanueva-Rosales, N.: Towards pharmacogenomics knowledge discovery with the semantic web. Brief Bioinform 10(2), pp. 153-163 (2009)
21. Marshall, M. S., Boyce, R., Deus, H. F., Zhao, J., Willighagen, E. L., Samwald, M., Pichler, E., Hajagos, J., Prudhommeaux, E., Stephens, S.: Emerging practices for mapping and linking life sciences data using RDF: A case series. Web Semantics: Science, Services and Agents on the World Wide Web 14(0), pp. 2-13 (2012)
22. Luciano, J., Andersson, B., Batchelor, C., Bodenreider, O., Clark, T., Denney, C., Domarew, C., Gambet, T., Harland, L., Jentzsch, A., Kashyap, V., Kos, P., Kozlovsky, J., Lebo, T., Marshall, S., McCusker, J., McGuinness, D., Ogbuji, C., Pichler, E., Powers, R., Prud'hommeaux, E., Samwald, M., Schriml, L., Tonellato, P., Whetzel, P., Zhao, J., Stephens, S., Dumontier, M.: The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside. J. Biomed. Semantics, 2 (Suppl 2), (2011)
23. Sim, I., Carini, S., Tu, S., Wynden, R., Pollock, B., Mollah, S., Gabriel, D., Hagler, H., Scheuermann, R., Lehmann, H., Wittkowski, K., Nahm, M., Bakken, S.: The human studies database project: federating human studies design data using the ontology of clinical research. In: AMIA Summits Transl. Sci. Proc. 2010, pp. 51-55 (2010)
24. Ogbuji, C.: A Framework Ontology for Computer-Based Patient Record Systems. In: 2nd Int. Conf. on Biomedical Ontology (ICBO-2011), CEUR-Proceedings, vol. 833, CEUR-WS.org (2011)
25. Kong, Y., Dahlke, C., Xiang, Q., Qian, Y., Karp, D., Scheuermann, R.: Toward an ontology-based framework for clinical research databases. J. Biomed. Inform. 44(1), pp. 48-58 (2011)
26. Laleci, G., Yuksel, M., Dogac, A.: Providing Semantic Interoperability between Clinical Care and Clinical Research Domains. IEEE trans. On Information Technology in Biomedicine, vol. PP(99), to appear (2012)
27. Neumayr, B., Schrefl, M., Linner, K.: Semantic Cockpit: An Ontology-Driven, Interactive Business Intelligence Tool for Comparative Data Analysis. In Olga De Troyer, Claudia Bauzer Medeiros, Roland Billen, Pierre Hallot, Alkis Simitsis, Hans Van Mingroot, ed.: Advances in Conceptual Modeling. Recent Developments and New Directions. Springer Berlin / Heidelberg, pp. 55-64 (2011)