A BFO-compliant Ontology for Radiation Therapy (RTO)

Mark H Phillips[1], Andre Dekker[2], and Jonathan Bona[3]

[1]Department of Radiation Oncology, University of Washington, Seattle, WA, USA
[2]Department of Radiation Oncology, University of Maastricht, MAASTRO, Maastricht, NL
[3]Department of Informatics, University of Arkansas, Little Rock, AR, USA

Abstract This work describes the development of an ontology for the medical domain of radiation therapy. This effort was undertaken to provide a solid foundation upon which other standards within the field can be developed and applied. Given the uncertainties in vocabularies and concepts, the decision was made to harness the W3C ontology tools which provides numerous advantages. These include interoperability between domains, a serialization that is widely acceptable and computable, and an implementation that allows the use of Descriptive Logic. The ontology follows the guidelines of the OBO Foundry and is based on the Basic Formal Ontology. It was written in the OWL language using the software package, Protege. Examples of the utility of this ontology in the clinical environment of radiation oncology are provided.

1 Introduction

The Open Biological and Biomedical Ontology Foundry (OBO Foundry) has emerged as the home of a large number of ontologies. To be included in this repository, the creators of an ontology must follow certain principles and best practices for creating and maintaining ontologies [1]. The framework established by the OBO Foundry serves a number of useful goals. Perhaps most importantly, it guarantees interoperability between ontologies. In addition, the use of a common format reduces the difficulties in using parts or all of an ontology. Finally, the use of unique IRI's and the commitment to avoiding identical labels of ontology entities insures that ambiguities are minimized.

The topics of the OBO Foundry ontologies span a very wide range, including three "base" ontologies: Basic Formal Ontology (BFO), Core Ontology for Biology and Biomedicine (COB), and Relation Ontology (RO). These greatly help in making sure that ontologies as separate as the Gene Ontology (GO) and Ontology for General Medical Science (OGMS) can use classes and relations from each other. Many of the ontologies are basic biology-oriented, while others serve as repositories of general domain knowledge.

In this work, we described the construction of an OBO-compliant ontology for the domain of radiation therapy (RTO). The motivation for this work is several-fold. First, several technologies are emerging that require, or at least benefit from, a comprehensive knowledge representation of this domain. These include initiatives such as mCODE [2] and IHE-RO [3]. Second, the field of radiation oncology suffers from a plethora of database schemas that include several different Oncology Information Systems (OIS), electronic health record systems (EHR), insurance companies, and national and international research organizations. In the absence of a standard representation of the field, these computer systems either cannot communicate or require extensive, and ever-evolving, software to map variables and concepts. Even that environment is fraught with potential errors since it is difficult to resolve ambiguities and/or redundancies. Finally, the rise of machine learning has highlighted the need for comprehensive structured data that is intelligible across medical and biological domains. For example, the potential to find associations between genetic patterns and oncology treatment outcomes requires a common knowledge representation. Even between medical subfields, such as medical, surgical and radiation oncology, unambiguous and explicit definitions are important.

It should be noted that the RTO takes as its domain "radiation therapy" which is not identical with the larger domain of "radiation oncology". The latter includes a much more comprehensive area of knowledge that is more biology-oriented than radiation therapy, which is more oriented towards the physics/engineering and clinical side of radiation oncology. If an ontology of radiation oncology and/or oncology in general is constructed, it is our expectation that the RTO will be easily integrated, thereby reducing the effort and possible duplication and confusion. Of course, it is not possible to cleanly separate radiation oncology from radiation therapy, and RTO does attempt to make bridges when it is necessary.

2 Methods and Materials

This project was implemented by the American Association of Physicists in Medicine (AAPM) as a task group charged with building an OBO-compliant ontology. (Simultaneously, the AAPM Big Data Subcommittee also began working on a non-compliant ontology in conjunction with the mCODE project; this ontology serves as a foundation for future work harnessing the Semantic Web [4].) Task group membership consisted of AAPM volunteers. In addition, an informatician (J Bona) with experience with OBO Foundry ontologies was recruited. Work on the project was carried out using a github repository and with biweekly meetings carried out over the internet. In addition, discussions with other OBO ontology groups were conducted to resolve issues.

The ontology construction was carried out using the standards promulgated by the OBO Foundry organization. The ontology was also built using the standards and tools of the World Wide Web (W3C) consortium [5]. This was done to insure the maximum flexibility and applicability of the ontology, as well as being able to leverage the software and tools developed under these standards.

The Ontology Development Kit (ODK) was used to initialize the ontology repository. The software package Protege [6] was used to develop the ontology using the W3C language OWL: "OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit." [7]

Our ontology was built upon the Basic Formal Ontology [8, 9] and the Relation Ontology [10]. These ontologies define the taxonomy of the ontology and the relations between individuals, respectively. They are extremely general, defining entities such as "process" and relations such has "has-part", and do not refer to any concepts specific to biomedical domains. In addition, the evolving standard Core Ontology for Biology and Biomedicine [11] was integrated when possible. This arose as it became clear that there were a number of classes that were so general to the domains of biomedicine, e.g. device, that it was useful to include them in a standard ontology which all interested users could import.

The Task Group relied on the tools built into Github to provide a controlled process for the discussion of issues, the addition to and modification of the ontology as it developed, and convenient means of coordinating between members of the task group as well as interested parties.

The knowledge representation of the domain of radiation therapy can be broken up into overlapping subdomains. This is due to the large degree of overlap between disparate fields such as nuclear physics, medical practice, oncology, computer engineering, and biomedical informatics. One of the main challenges in building this ontology is incorporating ontological bridges between these different domains. Given the OBO principles of re-use of terms within the collective ontologies, this led to the need of reviewing a large number of related ontologies, and when necessary, discussing and collaborating with these groups.

Although the subdomains were not completely orthogonal, we were able to sequentially focus on the following areas.

- basic and nuclear physics;
- production and measurement of ionizing radiation;
- radiation therapy workflow;
- clinical devices specific to radiation therapy.

Finally, an important component of the development of the ontology was the recognition that clinical practice evolves, new technologies are developed and terminology is difficult to control. Therefore, this ontology was built to contain as little of current vocabularies–rather, it tries to encapsulate the most basic concepts with well-defined terms and definitions. More specific classes can be built using the Descriptive Logic methods that are an integral part of the OWL standard. It is expected that users can apply this formal ontology to their specific needs by using axioms to build specific complex classes from the elemental concepts in RTO.

3 Results

The Radiation Therapy Ontology (RTO) was built and published under the auspices of the OBO Foundry. This version of RTO is a backbone upon which further details and refinements can be made. The classes and properties that were implemented were done so with an eye to several practical applications–namely mirroring the variables used in the Operational Ontology of Radiation Oncology [4], and providing support for a multi-institutional research project on reducing errors in radiation oncology [12].

Following the dictum that classes defined in other ontologies should be used, rather than re-defined, classes were imported from a number of different ontologies, including Basic Formal Ontology, Relation Ontology, Core Ontology for Biology and Biomedicine, Information Artifact Ontology, Radiation Biology Ontology, Chemical Entities of Biological Interest, Ontology for Biomedical Investigations, and Ontology for General Medical Science. In addition, the NCI Thesaurus is sometimes linked to definitions.

The goal of avoiding overly specific classes and vocabulary was achieved to some degree by making use of the concept of defined classes. For example, the concept of Intensity modulated radiotherapy (IMRT) means different things to different stakeholders. As a simplified example, a user could define a class `IMRT` as a subclass of `techinque action specification` and then restrict it as follows:

`has Part min 1 inverse_planned_specification`
and `has Part min 1 multileaf_collimator`
and `has Part min 5 control_point`

where the objects of these three triples are classes already defined.

The other goal was to interface with other clinical/research systems in radiation oncology. For this, we relied on Ontop [13]. Ontop is a Virtual Knowledge Graph system. It exposes the content of arbitrary relational databases as knowledge graphs. These graphs are virtual, which means that data remains in the data sources instead of being moved to another database. Ontop translates SPARQL queries (opens new window) expressed over the knowledge graphs into SQL queries executed by the relational data sources and relies on R2RML mappings. Again, as with the defined classes, this approach allows a maximum of flexibility between established data sources and emerging standards. R2RML has a default mapping scheme or it can be individually tailored as needed.

4 Discussion

The development of a radiation therapy ontology has become necessary in order for the field to be able to leverage the advances made machine learning as well as biological advances in other biomedical fields. The AAPM has been at the forefront of this effort with a two-pronged approach. First, it has worked with other stakeholders, such as American Society of Clinical Oncology and IHE-RO and HL7 FHIR, to construct an ontology based on the commercial and clinical environment [4]. The second is to work to apply the advantages and power of the Semantic Web to data curation by the creation of a formal ontology, RTO.

One of the persistent problems in making the most of the data we are continually acquiring is the set of conflicting standards, particularly in vocabulary. For example, the term "modality" is a term of art that is commonly used but not clearly defined. DICOM-RT Second Generation uses the term in multiple places with varying definitions depending on the context. The term is also defined in the Operational Ontology for Oncology. The standard defines a value set, none of which matches the DICOM standard. This leads to significant difficulties when trying to query relational databases.

There are several different approaches one can take with this ontology. If one is using Ontop, then R2RML can be used to map the RTO class to the corresponding appropriate column variable in the database. Alternatively, one can add a class specific to the application and then use the equivalence axiom to map one to the other.

As discussed earlier, the use of standards promulgated by the W3 Consortium provides a wealth of possibilities for expanded and novel uses of an ontology in radiation oncology. These include the ability to tap into the expertise of related domains, the reduction in the effort needed to broadcast data from a given domain, and the ability to leverage algorithms based on knowledge graphs and descriptive logic [14–16].

The representation of data in this approach differs from traditional relational databases. The core concept is the "triple"– subject, predicate, object. For example, `J Smith has_role patient`.

Figure 1 represents the way in which the overall concepts of the ontology are modelled for RT. In the database, each of the classes shown in Figure 1 would be instantiated by individuals, represented by an Internationalized Resource Identifier (IRI) which is unique. That is, similar to DICOM UID's, a given `radiation therapy plan specification` for a given patient would have a unique IRI which provides an unambiguous link to the resource.

An important aspect of the approach we have taken to building an ontology is the ability to use inference. The use of a defined syntax and semantics (in this case, RDF/XML) allows one to build a "reasoner" which is software that applies Desciption Logic to infer new facts from the stated ones using the axioms of the ontology. Simple examples include inferences such as "If A is a spouse of B, then B is a spouse of A", "If C is the child of D, D is the parent of C", and "If Patient X is treated on Y therapy device, and Y is a gamma-emitting device, then X is treated using gamma rays." In the latter example, the reasoner builds the latter triple using only

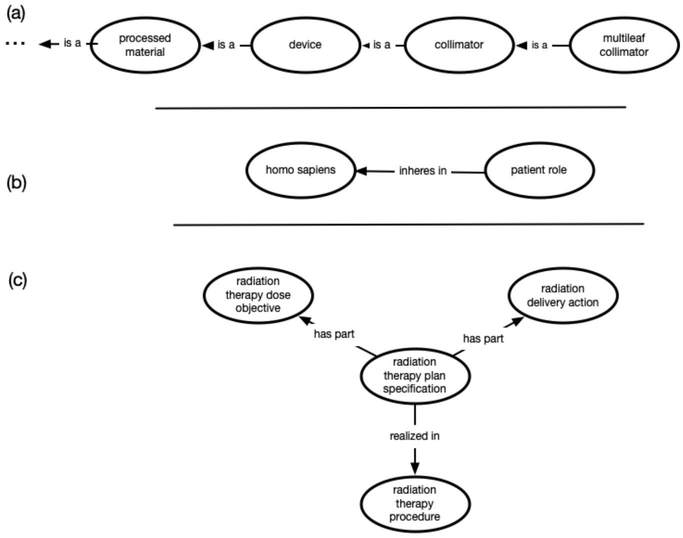

Figure 1: Elements of the ontology expressed as triples. For reasons of space, elements are rendered separately. (a) taxonomy representing a part of a linac; (b) representation of humans as patients; (c) elements that go into defining the use of radiation therapy as a treatment.

the first two without explicit construction.

As discussed earlier, this ability to use inference allows a specific application to define a useful class even though the available data does not explicitly contain that structured data. In this way, the formal RTO ontology can be used to provide building blocks for a wide range of applications and, with relatively minor additions, be tailored for the specific purposes.

The implementation of data using OWL has other advantages as well. For one, the ability of the reasoner to use inference makes it easy to do complex searches that would require difficult SQL joins. For example, searching a database for patients that met a complex set of eligibility criteria for a clinical trial can be simpler, depending on the criteria and size of the database. If one is going to store the data in the triple store approach, it can also be much easier to modify the schema as knowledge evolves and progresses relative to the corresponding effort with relational databases.

Finally, by designing RTO as an OBO-compliant ontology, progress in knowledge in related fields, such as genetics, can more easily be integrated into research and clinical practice of radiation oncology. In this way, the Semantic Web can provide computational tools to explore and understand how to better use radiation to treat cancer.

5 Conclusion

The motivation, methods and advantages of building a formal ontology following OBO Foundry priniciples for the radiation therapy domain has been described. This work was not meant to reduce the effectiveness of other approaches. Rather, it has been carried out in anticipation of its utility in the future development of computational methods utilizing a wide range of biological and biomedical data. Using OWL and Protege is a reflection that sometimes the model is as important or more important than the data whereas the relational database is useful when data are more important than the model.

References

[1] *Open Biological and Biomedical Ontology Foundry*. https://obofoundry.org/principles/fp-000-summary.html. Accessed 20 Jan 2024.

[2] *minimal Code Oncology Data ELements*. http://hl7.org/fhir/us/mcode/ImplementationGuide/hl7.fhir.us.mcode. Accessed 20 Jan 2024.

[3] *Integrating the Healthcare Enterprise – Radiation Oncology*. https://www.ihe.net/uploadedFiles/Documents/Radiation Oncology/IHE-RO-TF-Vol1.pdf. Accessed 20 Jan 2024.

[4] C. S. Mayo, M. U. Feng, K. K. Brock, et al. "Operational Ontology for Oncology (O3): A professional society-based, multi-stakeholder, consensus-driven informatics standard supporting clinical and research use of real-world data from patients treated for cancer". *International Journal of Radiation Oncology* Biology* Physics* 117.3 (2023), pp. 533–550.

[5] *World Wide Web Consortium*. https: www.w3.org. Accessed 25 Jan 2024.

[6] M. Musen. "The Protege Project: a look back and a look forward". *Association of Computing Machinery Specific Interest Group in Artificial Intelligence* 1 (4 2015). http://protege.stanford.edu.

[7] *"Web Ontology Language (OWL)*. https://www.w3.org/OWL/. Accessed 21 Jan 2024.

[8] *Basic Formal Ontology*. https://basic-formal-ontology.org. Accessed @3 Jan 2024.

[9] *Building Ontologies with Basic Formal Ontology*. MIT Press, 2015.

[10] *OBO Foundry: Relation Ontology*. http://obofoundry.org/ontology/ro.html. Accessed 23 Jan 2024.

[11] *OBO Foundry: Core Ontology for Biology and Biomedicine*. http://obofoundry.org/ontology/cob.html. Accessed 23 Jan 2024.

[12] P. Kalendralis, S. M. Luk, R. Canters, et al. "Automatic quality assurance of radiotherapy treatment plans using Bayesian networks: A multi-institutional study". *Frontiers in Oncology* 13 (2023), p. 1099994.

[13] *Ontop*. https://ontop-vkg.org/guide/#versions. Accessed 26 Jan 2024.

[14] J. Sanjak, J. Binder, A. S. Yadaw, et al. "Clustering rare diseases within an ontology-enriched knowledge graph". *Journal of the American Medical Informatics Association* 31.1 (2024), pp. 154–164.

[15] D. Kim, J.-G. Joung, K.-A. Sohn, et al. "Knowledge boosting: a graph-based integration approach with multi-omics data and genomic knowledge for cancer clinical outcome prediction". *Journal of the American Medical Informatics Association* 22.1 (2015), pp. 109–120.

[16] S. S. Hasan, D. Rivera, X.-C. Wu, et al. "Knowledge graph-enabled cancer data analytics". *IEEE journal of biomedical and health informatics* 24.7 (2020), pp. 1952–1967.