A semantic database for integrated management of image and dosimetric data in low radiation dose research in medical imaging

Bernard Gibaud, PhD[1], Marine Brenet, MSc[1], Guillaume Pasquier, MSc[2], Alex Vergara Gil, MSc[3,4], Manuel Bardiès, PhD[3,4], John Stratakis, PhD[5], John Damilakis, PhD[5], Nicolas Van Dooren, MSc[6], Joël Spaltenstein, MD, MSc[6], Osman Ratib, MD, PhD[6]

[1]Univ Rennes, Inserm, LTSI UMR 1099, Rennes, France
[2]B-COM Institute of Research and Technology, Rennes, France
[3]Centre Recherche en Cancérologie de Toulouse, Toulouse, France
[4]UMR 1037, INSERM, Université Toulouse III Paul Sabatier, Toulouse, France
[5]Medical Physics Department, School of Medicine, University of Crete, Heraklion, Greece
[6]Institute of Translational Molecular Imaging, Genève

Abstract

Medical ionizing radiation procedures and especially medical imaging are a non negligible source of exposure to patients. Whereas the biological effects of high absorbed doses are relatively well known, the effects of low absorbed doses are still debated. This work presents the development of a computer platform called Image and Radiation Dose BioBank (IRDBB) to manage research data produced in the context of the MEDIRAD project, a European project focusing on research on low doses in the context of medical procedures. More precisely, the paper describes a semantic database linking dosimetric data (such as absorbed doses to organs) to the images corresponding to X-rays exposure (such as CT images) or scintigraphic images (such as SPECT or PET images) that allow measuring the distribution of a radiopharmaceutical.
The main contributions of this work are: 1) the implementation of the semantic database of the IRDBB system and 2) an ontology called OntoMEDIRAD covering the domain of discourse involved in MEDIRAD research data, especially many concepts from the DICOM standard modelled according to a realist approach.

Keywords

Semantic technologies, Ontologies, Medical imaging, Imaging biobanks, Radiation protection, DICOM

Introduction

Radiation exposure from diagnostic medical imaging is sometimes significant and is known to present a risk of damage for cells and DNA. For this reason, radiation protection guidance and legislation apply a precaution principle (ALARA concept: as low as reasonably achievable) consisting in delivering the minimal irradiation compatible with the diagnostic procedure. However, the health implications of low to moderate exposure are still a subject of debate and more research is needed to get a better understanding of biological effects of ionizing radiations, of repair/regeneration mechanisms, but also of the relation between absorbed dose and image quality. All these topics are addressed in a wide EU project called MEDIRAD (Title: "Implications of Medical Low Dose Radiation Exposure"), in the EURATOM program (addressing the domain of research of low doses). In the context of this project, several clinical research studies are carried out, which involve medical procedures associated with exposure of patients to ionizing radiation, and for which the calculation of the absorbed doses in organs is performed.

This work addresses the development of a computer system called *Image and Radiation Dose BioBank* (IRDBB) designed to manage image and dosimetric data in an integrated way. More precisely, this article describes the components of this system that populate, store and query a semantic database facilitating the integrated management of image and dosimetric data. This semantic database is implemented as a Resource Description Framework (RDF) graph aligned onto an application ontology called OntoMEDIRAD, that specifies the semantics of any information within this database. This approach follows the general methodology proposed in [1].

Although the IRDBB system will be primarily used to fulfil the needs of the researchers involved in the MEDIRAD project, both the IRDBB system and the OntoMEDIRAD ontology were developed with an objective of extensibility

and reusability in the context of similar projects. The choice of an ontology-based approach aims eventually at facilitating the access to MEDIRAD research data to a wide community of researchers interested in low dose research, e.g. via federated systems.

The following of the paper is organized as follows. The 'Related works' section situates this contribution with respect to the state of the art. The 'Material and methods' section provides a description of the overall IRDBB system architecture. It explains how the OntoMEDIRAD ontology was built and how it is used in the component of the system called *Semantic Translator* to populate the semantic database. The 'Results' section provides details about the current content of the ontology and the domain covered. The 'Discussion' highlights some of the key choices we made in both the design of the ontology and the implementation of the Semantic Translator. It also reviews the main limitations of this work and situates it with regards to the state of the art. The 'Conclusion and perspectives' section emphasizes the main originality of the work and provides some hints for the continuation of the project.

Related works

The use of semantic technologies has been very successful in the domains of biology and life sciences. Ontologies such as Gene Ontology (GO) have been used universally for more than 15 years to annotate experimental data and scientific literature in both the omics and clinical realms. Large repositories of biological and bioinformatics data exist such as the European Bioinformatics Institute (EBI) RDF Platform that makes extensive use of RDF data and ontological resources [2].

In the domain of medical imaging, the use of such technologies is not so widespread yet. A review paper published in 2015 allowed inventorying the main ontologies and related initiatives worldwide in medical imaging, encompassing both radiology and histopathology [3]. Most works concern the management of the workflow in radiology [4], the annotation of medical images, and the "ontologization" of the Digital Imaging and Communications in Medicine (DICOM) standard [5], primarily applied to the domain of cancer research and radiation oncology [6]. The use in medical reasoning was investigated in the Theseus MEDICO project [7], but with limited actual deployment. Another active domain is neuroimaging, with the development of the NeuroImaging Data Model (NI-DM model) focusing on experimental design and study data. NI-DM makes use of semantic web technologies to describe information about the design and intent of neuroimaging experiments, subjects' characteristics, and acquired data, e.g. in the context of fMRI [8]. NI-DM terminologies are available, and also exist in OWL format[1], but they primarily aim at managing provenance and do not rely on any upper-level ontologies, limitations that compromise their use in complex multi-disciplinary application contexts. To our best knowledge, no developments have addressed the domain of patient radiation protection in medical imaging.

Material and methods

Basic system functionality

The overall IRDBB system is a platform designed: 1) to support the importation of research data sent by the MEDIRAD users, 2) to store this data, and 3) to provide tools enabling MEDIRAD users to query and retrieve this data. Basically, two kinds of data are involved: image data and dosimetric data. Image data are images corresponding to the exposure to ionizing radiation, e.g. chest computed tomography (CT) images in children and adults, or images acquired as part of a targeted radiotherapy procedure selected in the MEDIRAD project, namely ^{131}I treatment of differenciated thyroid cancer. The latter procedure requires performing several nuclear medicine (NM) explorations for locating precisely the distribution of the radiopharmaceutical in the body, a prerequisite for the calculation of absorbed doses to organs. Dosimetric data can be provided in various ways, either produced by imaging devices such as CT Radiation Dose Structured Reports (CT SR), or as results of calculations made by MEDIRAD researchers using advanced Monte Carlo based dosimetry methods. The data can be represented either in DICOM format or in some other non-DICOM format.

[1] https://github.com/incf-nidash/nidm-terms

IRDBB system architecture

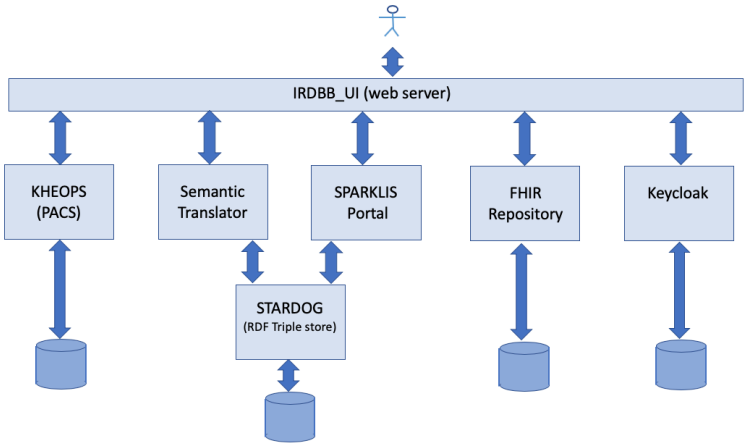


Figure 1. IRDBB system architecture

The overall architecture of the IRDBB system is shown in Figure 1. The major components are: 1) a component called *IRDBB_UI*, which is a web server managing the user interface; 2) a component called *KHEOPS,* managing the DICOM data (based on the DCM4CHE software); 3) a component called *FHIR repository,* managing all non-DICOM files; 4) a component called *Semantic Translator*, providing a set of services to populate and query the semantic database; 5) a *STARDOG* Triple Store, supporting the semantic database, 6) a component called *Sparklis Portal,* extending the IRDBB_UI to assist the users in building SPARQL queries, and 7) a component called *Keycloak,* providing a Single Sign-On mechanism for access control.

Design of the ontology

The overall semantic system was designed based on requirements collected at the beginning of the project by means of a questionnaire sent to all MEDIRAD users. The answers received allowed to specify which DICOM objects should be supported – Information Object Definitions (IODs) and corresponding Service Object Pair Classes (SOP classes), as well as what important metadata. These specifications were complemented in the course of the project with descriptions of the workflows allowing to produce the MEDIRAD research data, especially absorbed doses in organs calculated by MEDIRAD researchers, and the related provenance data (i.e. how a particular dataset was produced, using what process, what inputs, what method and method settings). All these specifications allowed to delimit the universe of discourse to be covered by the ontology.

The ontology was designed as an application ontology gathering all entities and relationships involved. The general modelling approach that we adopted was a realist one [9], i.e. trying to refer to entities existing in reality, rather than on conceptual constructs, and trying also to identify properly those real-world entities, as recommended in [10].

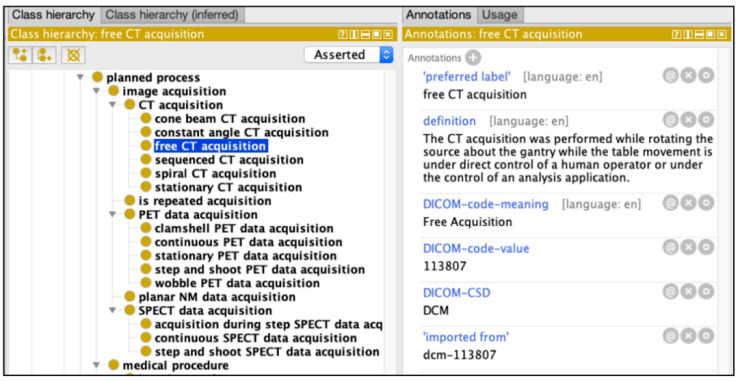
Of course, we tried as much as possible to reuse existing ontological resources. Therefore, we adopted an organization in modules, in which the root application ontology (called OntoMEDIRAD) imports several extracts of existing ontologies. These extracts, e.g. from the Foundational Model of Anatomy, the Units Ontology (UO), the Phenotype and Trait Ontology (PATO) were generated using the OntoFox tool[2] [11], based on the MIREOT model [12]. The overall integration of these disparate ingredients relied on the common philosophical ground provided by the Basic Formal Ontology (BFO version 2[3]) [13]. Of course, every entity involved in the application domain could not be found in existing ontologies. For example, we didn't find such resources for the domain of radiation protection. Consequently, we had to create the ontology classes, object properties and data properties concerning non covered parts of the domain of discourse. Concerning DICOM data, we relied on the entities considered in the DICOM terminology (Part 16). They have various origins, primarily SNOMED CT, but also the DICOM DCM terminology resource. Regarding the latter, we relied on existing IRIs available from the DICOM DCM ontology[4], available from the National Center of Biomedical Ontology (NCBO) Bioportal, and maintained by the editor of the DICOM Standard. This ontology is available in OWL but organized as a flat list of terms, so we had to re-organize these terms to integrate

[2] http://ontofox.hegroup.org/
[3] https://basic-formal-ontology.org/
[4] https://bioportal.bioontology.org/ontologies/DCM

them properly into our subsumption hierarchy. Another aspect of this modeling work was the analysis of DICOM Context Groups referred to in the DICOM Structured Reports templates that we needed to support, especially those related to CT SR dose reports. Context Groups provide flat lists of terms that can be used in a particular context in order to assign a coded value to a particular coded entry. All the items present in each Context Group had to be considered, and potentially dispatched in various places in the overall ontology taxonomy.

Coding Scheme Designator	Code Value	Code Meaning
DCM	113804	Sequenced Acquisition
SCT	116152004	Spiral Acquisition
DCM	113805	Constant Angle Acquisition
DCM	113806	Stationary Acquisition
DCM	113807	Free Acquisition
SCT	702569007	Cone Beam Acquisition

A - Table CID 10013. CT Acquisition Type (DICOM Part 16)



B – Corresponding part in the OntoMEDIRAD ontology

Figure 2. DICOM Context Groups: illustrative example;
Left part (A): Example of DICOM Context Group ;
Right part (B) Corresponding entities in the OntoMEDIRAD taxonomy

An illustrative example is presented Figure 2. In this example, Context Group CID 10013 is referred to in the SR Template TID 10013 CT Irradiation Event Data, in order to precisely specify the detailed characteristics of CT acquisitions. We created corresponding classes in the OntoMEDIRAD ontology, subsumed by a CT acquisition Class, that corresponds to the title of this Context Group. Note that, for DCM terms we kept original IRIs from the DCM ontology, e.g. http://dicom.nema.org/resources/ontology/DCM/113807. Besides, we kept the original DICOM Code Value and Code Meaning in the DICOM-code-value and DICOM-code-meaning Annotation properties, but we adopted a different (more explicit) label in our skos:prefLabel Annotation property, e.g. 'free CT acquisition' in this particular case.

Semantic Translator

This component provides several services to manage semantic data. The main service translates into RDF the metadata associated to the data files: once the data files have been imported into the corresponding repository (i.e. KHEOPS repository for DICOM data and FHIR repository for non-DICOM data) the associated key metadata are translated into RDF to populate the semantic database. This process involves creating instances of the classes of the ontology, and assigning corresponding values by means of data properties available in the ontology (for example the date and time of a particular CT acquisition). It also consists in connecting together instances using object properties of the ontology (for example to denote that a particular CT image was an output of a particular CT acquisition). The metadata had two major origins. For DICOM data files, they were retrieved from DICOM data elements present in the DICOM SOP instances, and characterizing, e.g., the Patient, the Study, the Series or the particular Image at stake, e.g. a CT image or a NM image. For DICOM Structured reports such as CT SR, information was also retrieved from the SR content tree, a complex structure embedded in DICOM elements that are compliant with corresponding SR Templates documented in DICOM Part 16.

For Non DICOM files, the metadata was retrieved from an XML file associated to the File Set imported by the user. This XML file contains references to all the files of the File Set that need to imported, as well as detailed information about each of them, namely: its nature, format, the related patient, the related MEDIRAD clinical research study, and the detailed file provenance (i.e. how it was produced, from which input data etc.). This XML file must comply to a predefined structure specified in an XML schema, whose items are aligned with the OntoMEDIRAD ontology.

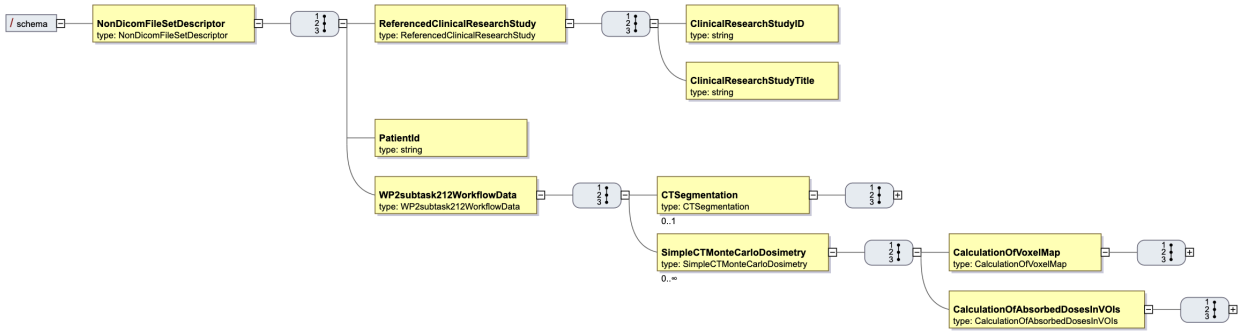


Figure 3. Extract of the XSD file describing the workflow of a simple Monte Carlo dosimetry (Blocks appearing with a '+' are themselves composed of blocks not shown on the figure).

Figure 3 shows an extract of the XML schema describing the workflow of a simple Monte Carlo dosimetry, consisting in calculating absorbed doses in organs or tissues delineated from CT images. The ReferencedClinicalResearchStudy block contains information identifying the research study in the context of which this calculation was made; the PatientId allows identifying the patient; the WP2subtask212WorkflowData block contains two blocks that further detail the two main steps of the calculation, namely the CTSegmentation, describing the segmentation of volumes of interest (VOI) and SimpleCTMonteCarloDosimetry introducing the two main steps, further described in the CalculationOfVoxelMap and the CalculationOfAbsorbedDosesInVOIs blocks.

All RDF triplets corresponding to the importation of a series of DICOM images or to a Non-DICOM File Set are serialized into an RDF graph, that is appended to the semantic database, supported by the STARDOG Triple Store. As a result, the semantic database is currently made of a single RDF graph, that includes both the ontology files, and all the elementary RDF graphs produced by the Semantic Translator upon the upload of DICOM and non-DICOM data.

Other services of the Semantic Translator concern the querying of the semantic database. As for retrieving detailed descriptions of image or dosimetry data files, specific SPARQL queries have been prepared, that can be selected by the user with the IRDBB_UI interface, and forwarded to the STARDOG server through the Semantic Translator query service. The result of the query is managed by the IRDBB_UI for presentation to the end user and export into CSV files.
The Semantic Translator was developed in Java.

Sparklis
The Sparklis component is an HTML5/Javascript application allowing the user to build SPARQL queries interactively. This module was developed by Sébastien Ferré in IRISA, Rennes (France) [14]. It was deployed as a complementary module of IRDBB_UI, extending the query service provided by the Semantic Translator. It allows the efficient building of SPARQL queries by pre-exploring both the ontology (i.e. the ontology classes and properties) and the actual data through a SPARQL endpoint. Thus, it allows the user to easily focus on those classes, object properties or

data properties for which values or instances actually exist in the RDF graph. It can be used by users with no particular knowledge of the SPARQL syntax, and it doesn't require detailed knowledge of the data schema.

Results

The main characteristics of the OntoMEDIRAD ontology are shown in Table 1. All modules are represented in OWL. It is freely available for consultation and download[5].

Module	Origin	No of classes
OntoMEDIRAD (Root module)	(MEDIRAD project)	606
FMA extract	Foundational Model of Anatomy (FMA)	362
PATO extract	Phenotype And Trait Ontology (PATO)	42
UO extract	Units Ontology (UO)	55
MPATH extract	Mouse Pathology (PATO)	4
ChEBI extract	Chemical Entities of Biological Interest (ChEBI)	81
Radionuclides extract	SNMI (SNOMED CT)	113
Radiopharmaceuticals extract	SNMI (SNOMED CT)	121
IAO extract	Information Artefacts Ontology (IAO)	11
Basic Formal Ontology (BFO)	Basic Formal Ontology - Version 2 (BFO)	37

Table 1. Characteristics of the OntoMEDIRAD Ontology

The Semantic Translator currently supports those DICOM entities that were involved in the clinical research studies of MEDIRAD whose research data had to be managed in the IRDBB system, namely CT images, PET Images, NM images, CT SR Dose Reports and Enhanced Structured Reports. The latter DICOM entity was used to represent Case Report Forms related to a particular clinical research study focusing on ^{131}I targeted radionuclide therapy (TRT) of differentiated thyroid cancer (DTC). The list of corresponding DICOM IODs and SOP classes is provided in Table 2.

DICOM Service Object Pair (SOP) Class	DICOM Information Object Definition (IOD)
CT Image Storage SOP Class	CT Image IOD
Enhanced CT Image Storage SOP Class	Enhanced CT Image IOD
Positron Emission Tomography Image Storage SOP Class	Positron Emission Tomography Image IOD
Enhanced PET Image Storage SOP Class	Enhanced PET Image IOD
Nuclear Medicine Image Storage SOP Class	Nuclear Medicine Image IOD
Enhanced Structured Reporting Storage SOP Class	Enhanced Structured Reporting IOD
X-Ray Radiation Dose SR storage SOP Class	CT Radiation Dose SR IOD

Table 2. DICOM SOP Classes and IODs currently supported

Concerning the management of Non-DICOM data, five workflows are currently supported:

- Simple Monte Carlo based estimation of absorbed doses from CT
- 3D Dosimetry in ^{131}I TRT of DTC – 1^{rst} approach
- 3D Dosimetry in ^{131}I TRT of DTC – 2nd approach

[5] https://github.com/OsiriX-Foundation/MediradOnto

- 2D Dosimetry in ^{131}I TRT of DTC
- Hybrid (i.e. combined 2D and 3D) Dosimetry in ^{131}I TRT of DTC.

The first describes a simple calculation process, involving three main steps: 1) segmentation leading to the delineation of 3D volumes of interest (VOIs), 2) Monte Carlo simulation of a CT acquisition, allowing to calculate absorbed doses in each voxel of the human body (more precisely in the part of the body imaged within the field of view - FOV - of the CT images), 3) calculation of mean absorbed doses in the different VOIs.

The four latter workflows describe the calculation of absorbed doses in several organs or tissues of interest, after the administration of ^{131}I (NaI) in TRT of DTC, following a common scheme involving: 1) acquisition of NM images at several timepoints (e.g. five timepoints) allowing to characterize the distribution of the radioactive material in the body along time, 2) the calculation of absorbed doses or absorbed dose rates, and 3) time-integration allowing to estimate the absorbed doses in tissues and organs of interest. The various workflows correspond to variants of this calculation process, e.g. in the way NM images are acquired and used - in 3D (3D dosimetry), 2D (2D Dosimetry) or using both (Hybrid Dosimetry), or in the way the integration of time-varying parameters is dealt with, i.e. if time integration is performed on activity or absorbed dose rates.

Sparklis

Figure 4. provides an illustration of the creation of a SPARQL query using Sparklis. The translation into pseudo-natural language appears in the upper-left part of the window: in this example, it retrieves the SPECT data acquisitions that are part of an imaging study whose target anatomical region is the head. The query may be further modified, e.g. to retrieve the date and time of these SPECT acquisitions, or select the images that are the output of these acquisitions: such properties for which data exist in the RDF graph appear in the middle-left part of the window.

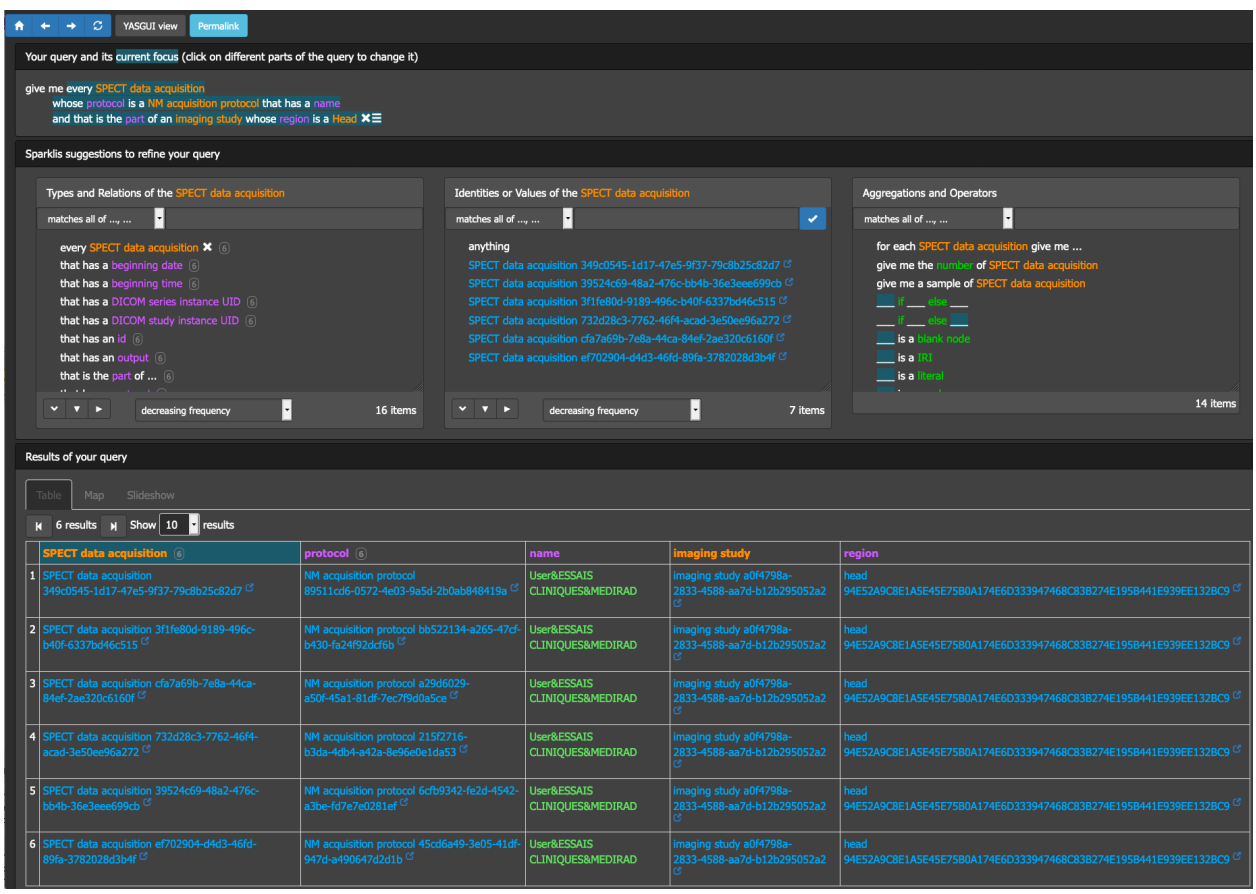


Figure 4. Example of query built with Sparklis

Discussion

In this work, our primary goal was to deploy a semantic database in the context of a professional platform dedicated to biomedical research in a pluri-disciplinary domain involving low doses delivered in medical imaging. This is one of the very first real-life deployments of semantic databases in the domain of imaging biobanks [15]. Imaging biobanks [16] will play a prominent role in future biomedical research, by facilitating the sharing and successful reuse of research data. They will also make it possible to provide developers of Artificial Intelligence algorithms with appropriate training data. There is no doubt that semantic technologies will be very beneficial in this context to properly categorize the data and manage the associated metadata, in a way that can satisfy the F.A.I.R. principles [17], i.e. actually make the data Findable, Accessible, Interoperable and Reusable.

This work demonstrates that even in the constrained context of a project such as MEDIRAD, it was possible to design the ontology, to develop the software for populating the semantic database and to provide the end users with appropriate tools to import, query and retrieve the data. Our work also demonstrates how semantic technologies can complement - rather than replace - traditional (e.g. relational) databases. For example, in the context of the IRDBB system, images can be retrieved directly from the KHEOPS DICOM server, and the semantic database brings a complementary, more integrated management, especially relating DICOM images to non-DICOM dosimetric data. In this respect, our work should help demystify semantic technologies in the medical imaging community.

The following of this discussion focuses on important aspects of our methodology, enhancing the capabilities and the limitations of our work.

Processing of semantic queries

When designing an ontology, a major issue is to choose the appropriate level of complexity. The richer the embedded knowledge, the more reasoning capabilities will be allowed. However, one must obviously also consider performance issues. In this regard, we adopted a rather conservative approach to guarantee efficient processing of SPARQL queries. In practice, we designed the pre-prepared SPARQL queries in such a way that they do not require complex reasoning (e.g. Description Logics reasoning) by the STARDOG engine. Moreover, OPTIONAL clauses in SPARQL queries were used with caution, especially to avoid dependencies between multiple OPTIONAL clauses, that may cause a dramatic increase of processing time.

This conservative approach was dictated by the need of providing rapidly a system offering adequate performance. Obviously, we are interested at exploring more ambitious uses of the knowledge embedded in the ontology, e.g. to automatically infer dependency graphs between image and absorbed dose data derived from the images, but this would need to implement suitable organization of the semantic database, e.g. by separating the RDF data into different graphs, or by using several versions of the ontology offering various expressivity.

Design of the ontology

As for DICOM, we analyzed in detail existing ontologies, especially the Semantic DICOM ontology[6] (SEDI) produced by SOHARD Software in Germany. Although it may cover the whole standard, we considered that it was not compatible with our realist modelling approach, since it very much relies on the specific internal organization of DICOM specifications (e.g. Information entities, Information Object Definitions, Sequences, Sequence Items) rather than on the nature and relations of the entities in the real world. Consequently, we applied our realist modelling approach to the DICOM entities falling in our application's scope, and we believe this was an interesting experience that should be generalized to the whole DICOM standard, even though this would represent a huge work, given the size and complexity of this standard. Nevertheless, we made our best efforts to stick to original DICOM semantics, e.g. by adopting original DICOM IRIs whenever existing and incorporating/adapting DICOM labels and definitions (retrieved from Part 3 and Part 16 of DICOM), especially in our annotation properties.

Semantic Translator

Translating into RDF the key DICOM metadata associated with DICOM objects was more complex than expected. A common and relatively simple approach was applied for all supported DICOM objects, consisting in instantiating the class of dataset and the class of process that produced this dataset, and relating them together using the *'has specified output'* object property, and associating to this process the various settings corresponding to the key elements of the imaging protocol, e.g. for CT: tube voltage, tube current, whether tube current modulation was used or not, etc. In practice, the main difficulties arose from three origins: 1) the fact that involved entities in the real world are not always well identified in DICOM headers (for example pieces of equipment, algorithms); 2) the fact that DICOM metadata

[6] https://bioportal.bioontology.org/ontologies/SEDI

sometimes includes several ways to describe the same information; for example, protocol settings can be found in single data elements or in complex sequences of items, thus requiring to explore all possible ways of encoding it; 3) the fact that semantically important information about the imaging procedure or the imaged organ is often encoded in free text or simply not present. All these constraints make it very difficult to formalize the DICOM to RDF translation process, at least in the relatively limited time frame that was available to us.

The management of non-DICOM data in the context of dosimetry was necessary because the related DICOM standard (namely the Patient Radiation Dose Structured Report) was not supported by the dosimetry applications developed by the MEDIRAD research groups. Consequently, it was necessary to specify several models of description of dosimetry workflows (modelled as XML schemas), suitable to describe the detailed provenance of dosimetric data, which was really challenging in the context of TRT because the dosimetry methodology is not yet standardized. As for TRT, such models were designed, taking into account existing recommendations of good practice of clinical dosimetry reporting [18]. They were implemented in the TRT dosimetry software developed at CRCT but not by the other groups yet. Furthermore, this dosimetry software was developed quite late in the project, thus requiring late and multiple extensions of the ontology and of the mapping of the ontology with the XSD schema constraining the structure of the XML data. Therefore, efficient tooling had to be developed to automate the creation of XSD files and produce and maintain the Java software responsible for translating the embedded information.

SPARQL queries

The complexity of the SPARQL query language is certainly one of the factors limiting the widespread diffusion of semantic technologies. For end-users, the difficulties arise from both the intrinsic complexity of the syntax and from the necessity to have a perfect knowledge of the data schema, i.e. the name and hierarchy of the classes, and the precise object properties and data properties used in the RDF graph. This is the reason why we made the choice of pre-prepared queries focusing on the main image and dosimetric datasets present in the semantic database. In the design of the queries we made the choice of broad queries (e.g. listing all CT images or all PET images), complemented by a filtering mechanism managed at the web browser level (e.g. to restrict the list to those concerning a particular patient or a particular clinical research study). An alternative could have been to ask the user to specify search criteria, which raised the difficulty of providing him/her with corresponding values. The advantage of such pre-prepared queries is that they can contain the whole set of acquisition parameters present in the database.

Nevertheless, it was obvious that a complementary mechanism had to be provided to enable a user to explore the graph in a flexible way. The integration of the Sparklis engine perfectly addressed this need. In practice, we realized that we had to provide a new set of labels for object properties and data properties, that is more consistent with the way Sparklis translates the SPARQL query in pseudo-natural language. This could be done very simply by adding a module in the ontology called SparklisLabels that provides these labels as additional annotation properties associated with each object property and data property. Moreover, it allowed introducing more simple labels, especially concerning properties provided in the BFO upper-level ontology, that might have been difficult to understand by the end-users.

We believe that motivated end-users will manage to use Sparklis, e.g. based on recorded videos providing concrete examples, but we have not sufficient return on experience, yet.

Conclusion and perspectives

The main contributions of this work are: 1) an implementation of a computer system called IRDBB including a semantic database to manage image and dosimetric data in an integrated way; this system has been intensively tested during the four last months and is now deployed in the ITMI facility in Geneva, and ready for use for supporting MEDIRAD clinical research studies. 2) An ontology called OntoMEDIRAD covering the domain of discourse involved in MEDIRAD research; this ontology models many concepts from the DICOM standard, according to a realist approach, which is really innovative since existing DICOM ontologies are either a flat list of terms or not designed according to a realist approach. This realization may help to promote the idea of designing a full-fledged DICOM ontology, which would significantly boost the development of imaging biobanks and AI software.

Acknowledgements

This work was supported by the European Commission as part of the MEDIRAD project (Number 755523) in the Horizon 2020 Program (EURATOM NFRP-2016-2017). The authors warmly thank all the MEDIRAD users who actively participated in the specification and the testing of the IRDBB system. The authors thank the STARDOG company for the free license of the STARDOG system, and for their efficient technical support. The authors thank

Sébastien Ferré (IRISA, Rennes) for providing a free license and technical assistance in the integration of the Sparklis system.

References

1. Neuhaus F, Vizedom A, Baclawski K, Bennett M, Dean M, Denny M, Grüninger M, Hashemi A, Longstreth T, Obrst L, Ray S, Sriram R, Schneider T, Vegetti M, West M, Yim P. Towards ontology evaluation across the life cycle: the ontology summit. Appl Ontol 2013; 8(3):179–194.
2. Jupp S, Bolleman J, Brandizi M, Davies M, Garcia L, Gaulton A, Gehant S, Laibe C, Redaschi N, Wimalaratne SM, Martin M, Le Novère N, Parkinson H, Birney E and Jenkinson AM. The EBI RDF platform: linked open data for the life sciences. Bioinformatics 2014;30(9)1338-1339.
3. Smith B, Arabandi S, Brochhausen M, Calhoun M, Ciccarese P, Doyle S, Gibaud B, Goldberg Ilya, Kahn CE Jr, Overton J, Tomaszewski J and Gurcan M. Biomedical imaging ontologies: A survey and proposal for future work. J Pathol Inform 2015;6:37.
4. Kahn CE, Channin DS and Rubin DL. An ontology for PACS integration. Journal of Digital Imaging 2006;19(4):316-327.
5. Van Soest J, Lustberg T, Grittner D, Marshall MS, Persoon L, Feltens P and Dekker A. Towards a semantic PACS: Using Semantic Web technology to represent imaging data. Study Health Technol Inform. 2014: 205:166-170.
6. Traverso A, Van Soest J, Wee L and Dekker A. The radiation oncology ontology (ROO): Publishing linked data in radiation oncology using semantic web and ontology techniques. Medical Physics 2018;45(10):e854-e862.
7. Oberkampf H, Zillner S, Bauer B and Hammon M. Interpreting Patient Data using Medical Background Knowledge. International Conference on Biomedical Ontologies (ICBO),2012;KR-MED Series, Graz, Austria.
8. Maumet C, Auer T, Bowring A, Chen G, Das S, Flandin G, Ghosh S, Glatard T, Gorgolewski K, Helmer KG, Jenkinson M, Keator DB, Nichols BN, Poline JB, Reynolds R, Sochat V, Turner J and Nichols TE. Sharing brain mapping statistical results with the neuroimaging data model. Scientific Data, Nature Publishing Group, 2016;3. DOI: 10.1038/sdata.2016.102
9. Smith B. From concepts to clinical reality: An essay on the benchmarking of biomedical terminologies. Journal of Biomedical Informatics 2006; 39, 288–298.
10. Ceusters W, Smith B. Strategies for referent tracking in electronic health records. Journal of Biomedical Informatics 2006; 39:362-378.
11. Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: web-based support for ontology reuse. BMC Research Notes. 2010; 3:175.
12. Courtot M, Gibson F, Lister AL, Malone J, Schober D, Brinkman RR, Ruttenberg A. MIREOT: the minimum information to reference an external ontology term. In: International conference on biomedical ontology, ICBO 2009; Buffalo, NY, USA.
13. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, The OBI Consortium, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 2007;25:1251–1255.
14. Ferré S. Sparklis: a SPARQL Endpoint Explorer for Expressive Question Answering. ISWC Posters & Demonstrations Track, Oct 2014; Riva del Garda, Italy. ⟨hal-01100319⟩
15. Hwang KH, Lee H, Koh G, Willrett D and Rubin DL. Building and Querying RDF/OWL Database of Semantically Annotated Nuclear Medicine Images. J Digit Imaging 2017;30,4–10. https://doi.org/10.1007/s10278-016-9916-7
16. SR Position paper on imaging biobank. Insights Imaging 2015;6(4):403-10, 2015. DOI 10.1007/s13244-015-0409-x
17. Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3, 160018 DOI 10.1038/sdata.2016.18
18. Lassmann M, Chiesa C, Flux G and Bardiès M. EANM Dosimetry Committee guidance document: good practice of clinical dosimetry reporting. Eur J Nucl Med Mol Imaging 2010; DOI 10.1007/s00259-010-1549-3.