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

S71 Semantics modeling and analysis

Bernard GIBAUD
Equipe MEDICIS, LTSI Inserm 1099, Université de Rennes 1
Disclosure

I have no relevant relationships with commercial interests to disclose.
Learning Objectives

After participating in this session the learner should be better able to:

- develop semantic databases, i.e. database whose data schema is aligned onto an application ontology
- build an application ontology, especially by integrating extracts of existing ontologies
- understand the DICOM standard and ways to represent DICOM entities in an ontology
A semantic database for integrated management of image and dosimetric data in low radiation dose research in medical imaging

Bernard Gibaud, PhD¹, Marine Brenet, MSc¹, Guillaume Pasquier, MSc², Alex Vergara Gil, MSc³,⁴ Manuel Bardiès, PhD³,⁴ John Stratakis, PhD⁵, John Damilakis, PhD⁵, Nicolas Van Dooren, MSc⁶, Joël Spaltenstein, MD, MSc⁶, Osman Ratib, MD, PhD⁶

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

• MEDIRAD project
  • EU project (EURATOM program), 2017-2022
  • 33 partners and 10 M€ EU funding
  • Project’s goal: The ultimate goal of MEDIRAD is to better understand and evaluate the health effects of low-dose ionising radiation exposure from diagnostic and therapeutic imaging and from off-target effects in radiotherapy.

• Task 2.4.1: Development of Image and Radiation Dose BioBank (IRDBB)
  • To manage the image and dose data involved in several MEDIRAD clinical projects
  • To create a software resource reusable for similar project
Objective

• To provide an integrated resource to manage images and dosimetric data

• Design principles: adherence to the F.A.I.R. principles, so that to make data use feasible: F(indable) A(ccessible) I(nteroperable) R(eusable)

especially through:
  • associating rich annotations to data files
  • associating explicit semantics to the annotations, by means on an ontology
  • relying on standards (DICOM), and existing ontologies, as much as possible
Needs analysis

• Questionnaire: Users were invited to describe
  • their research data (structure and format)
  • how this data should be shared and accessed (competency questions)

• Main functional needs
  • Calculation of absorbed doses (Chest CT, in both child and adult) – WP2
  • Calculation of absorbed doses (Hybrid imaging SPECT-CT and PET-CT) – WP2
  • Calculation of absorbed doses in organs and tissues in internal radiotherapy (treatment of thyroid cancer $^{131}$I) - WP3
  • Calculation of absorbed doses (pediatric CT) – WP5

• Data formats
  • Both DICOM and non-DICOM data
OntoMEDIRAD ontology*

* Available at: https://github.com/OsiriX-Foundation/MediradOnto/tree/master/ontology
Ontology

Available at: https://github.com/OsiriX-Foundation/MediradOnto/tree/master/ontology
Overall IRDBB Architecture

IRDBB_UI (web server)

- KHEOPS (PACS)
- Semantic Translator
- SPARKLIS Portal
- FHIR Repository
- Keycloak

STARDOG (RDF Triple store)
Semantic Translator

• Implemented in Java

• Library of services for accessing the semantic database, called by IRDBB_UI, especially:
  • Populating the semantic database at importation of DICOM or non-DICOM data
  • Executing predefined SPARQL queries
DICOM File Set

DICOM metadata

(0010,0020) PatientID

Patient

Image

series

study

Semantic Translator

Ontology

aligned onto

Semantic graph
- instances of the classes of the ontology
- data values (e.g. integer, float, date)
- relationships between instances

Ontology

DICOM File Set

DICOM metadata

(0010,0020) PatientID

Patient

Image

series

study
Semantic Translator

From dosimetry
- UoC (WP2)
- CRCT Toulouse (WP2, WP3)

XML File set Descriptor

Non-DICOM File Set

Data file

XML Schema

complies with

Ontology

aligned onto

Semantic graph
- instances of the classes of the ontology
- data values (e.g. integer, float, date)
- relationships between instances

Semantic Translator

Non-DICOM File Set

File set Descriptor

Data file

XML Schema

Ontology
Definition of the content of the XML File set descriptors (XML schema)
Semantic graph: total absorbed dose in organs
(Thyroïd $^{131}$I Targeted Radionuclide Therapy)
Querying the semantic graph

1. Predefined SPARQL queries
2. Navigating with SPARKLIS
Querying the graph with SPARKLIS

![Image of SPARKLIS interface]

**Your query and its current focus**

**Query:**
- give me every CT image dataset that is the output of a CT acquisition that a patient that is the role of a human is realized in and whose setting is a KVP that has a measurement value and that has a measurement unit and that has a DICOM series instance UID and that has a format.

**Results:**

<table>
<thead>
<tr>
<th><strong>CT image dataset</strong></th>
<th><strong>CT acquisition</strong></th>
<th><strong>patient</strong></th>
<th><strong>human</strong></th>
<th><strong>DICOM series instance UID</strong></th>
<th><strong>format</strong></th>
<th><strong>setting</strong></th>
<th><strong>measurement unit</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td>CT acquisition c2d4eab-0c8e-4cfe-83d6-61d997a23099</td>
<td>Patient 1537571c-9733-4d3c-9a80-45a6e0b3c3a6</td>
<td>Human 3b95ab44-662f-470a-9f3c-312b3e912c89</td>
<td>1.3.12.2.11075.1.4.50256.3000000711380641356200001249</td>
<td>Knee 4816065-564b-46d0-a976-420b219a62b2</td>
<td>120</td>
<td>KVP</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>CT image dataset 30a3e20a-b464-4263-b26b-256a64a23092</td>
<td>CT acquisition 4a56c6-4e8c-429d-bb63-fa11be5d6b1e</td>
<td>Human 3b95ab44-662f-470a-9f3c-312b3e912c89</td>
<td>1.3.12.2.11075.1.4.50256.3000000711380641356200001249</td>
<td>Knee 4816065-564b-46d0-a976-420b219a62b2</td>
<td>120</td>
<td>KVP</td>
</tr>
</tbody>
</table>
Current status

• IRDBB system intensely tested since mid-2019

• Data importation
  • DICOM data importation available since mid-2019
  • Non-DICOM data importation available since mid-2020 (WP2 and WP3 workflows for 3D dosimetry)

• Retrieval possible at both
  • the KHEOPS server level (for DICOM data)
  • the integrated IRDBB level (for any data, through the semantic database)

• IRDBB operational since September 2020
Discussion

• Why a semantic approach？
  • Semantic precision critical for data use and reuse in research
  • Best strategy w.r.t. F.A.I.R. principles
  • Definitely innovative in the domain of imaging biobanks
  • Context of relatively small repository to manage
  • Rich possibilities for extending the Semantic Translator with rule-based semantic services

• Main limitations
  • Functional: Need to specify consensual domain-specific data processing workflows
  • Technical: No general framework to manage RDF translation
Conclusion

• Innovative implementation of imaging biobank based on use of semantic graph

• Semantic web implementation complementary to standard image management approaches

• OntoMEDIRAD ontology, based on BFO, composed of several extracts, and integrating many DICOM terms
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 partners who participated in the specification and test of the system
  • Gordon Arnhold, Glenn Flux, Andreas Hinsch, Michael Lassmann, Peter Mildenerger, Lothar Schmidt, Tino Schurrat, Jan Tapperoge, Isabelle Thierry-Chef, Delphine Vallot

• And the STARDOG Company (for providing a free license)

Thank you for your attention!

Email me at: bernard.gibaud@univ-rennes1.fr