

Ontology-based classification of radiological procedures for consistent sharing in Clinical Data Warehouses

Pierre LEMORDANT^{a,b,1}, Bernard GIBAUD^a, Cyril GARDE^b, Sébastien DELARCHE^c, Didier GOUDET^d, Marc CUGGIA^a

^a *Univ Rennes, Inserm, LTSI UMR 1099, Rennes, France,*

^b *Enovacom, Marseille, France,*

^c *Centre Hospitalier Universitaire Pontchaillou, Rennes, France*

^d *Centre Eugène Marquis, Rennes, France*

Abstract. Clinical data warehouses (CDW) allow the reuse of care data in a research context. Designing and operating CDWs require addressing interoperability, data enrichment and data modeling problems, among others. This work concerns the management of medical imaging data in CDWs. It proposes a data-driven approach for classifying radiological procedures using an ontology-based approach. This approach relies on the RadLex ontology and an imaging procedures terminology called RadLex Playbook, both developed by RSNA. We first created an ontology of the radiological procedures by merging the Playbook with the relevant extract of the RadLex ontology and enriched it with French terms using the UMLS meta thesaurus. Then, we developed a proof of concept of a radiological procedures data classifier that exploits the richness of RadLex ontology and the ontological reasoning and we assessed it using medical imaging data retrieved from two different facilities. Our results demonstrate feasibility and relevance of the approach. They also highlight differences in the methods of filling imaging procedure data in the two institutions, as well as some problems in the RadLex ontology. Based on this experience, this proof of concept will be refined to evolve towards a routinely usable classification tool supporting medical imaging data management in CDWs.

Keywords. ontology, RadLex, playbook, clinical data warehouse,

1. Introduction

Clinical data warehouses (CDW) allow reusing care data for research purposes, and some inter-institutional projects also propose to gather data from several CDWs at regional or national levels. CDW gather data from different poles of the hospital and allow researchers to conduct their studies by providing them with tools to select patient cohorts, perform analyses or import these data into their own analysis tools. The main challenges to be addressed are therefore the heterogeneity of health data and interoperability between health care institutions. The “Massive Health Data” team of the LTSI Laboratory develops and maintains eHOP, a CDW technology within the Rennes University Hospital [1]. In order to improve our CDW’s versatility, we want to add to it

¹ Corresponding Author, Pierre Lemordant, Univ Rennes 1, Inserm, LTSI UMR 1099, Rennes, France; E-mail: pierre.lemordant@univ-rennes1.fr.

the ability to manage medical imaging information. Therefore, we plan to extract from the hospital's Picture Archiving and Communication Systems (PACS) the structured data that characterize imaging procedures and to use it in the hospital's datawarehouse, to index, classify and enrich this data. In the long run, we consider artificial intelligence or data mining study cases that will benefit from the alignment of our imaging data in a model describing in depth the entire field of radiology and accessible to algorithms. Ontologies provide a way to meet these needs and, in this work, we propose to categorize local imaging procedures using an ontology of radiology procedures.

The biomedical domain is a favorable ground for the development of ontologies. Many ontologies have been developed to describe different aspects of medicine : anatomy with FMA (Foundational Model of Anatomy), genomics with GO (Gene Ontology), radiation oncology with ROO (Radiation Oncology Ontology) as well as many others. The leading initiative to describe the field of radiology, which is of interest to us in this work, is RadLex [2][3]. The RadLex ontology has been produced by the [Radiological Society of North America \(RSNA\)](#) in order to provide radiologists with a common vocabulary. Started in 2005, RadLex now includes more than 30,000 terms and covers all fields of radiology like modalities (i.e. imaging techniques), anatomy, pharmacology and so on. In addition, the RSNA designed the RadLex Playbook [4], intended to be a terminology for radiology procedures. The current version of the Playbook is the result of a harmonization work between RadLex and the Logical Observation Identifiers Names & Codes (LOINC) terminology; it defines a list of radiological procedures, describing them through elements of the RadLex ontology. However, the procedures themselves are not integrated into RadLex.

Medical images and their related data are managed using the international DICOM standard [5]. This standard combines imaging data with relevant metadata such as patient demographics, examination information, device settings, etc. This is the format in which we receive the information about imaging procedures that we have to classify. Depending on the institution or modality manufacturer, DICOM metadata are more or less filled in, with information of variable quality. In our case our data come from two different institutes, which gives us the opportunity to make our solution more generic. DICOM defines a vocabulary for imaging metadata but does not define a standard terminology for procedures themselves.

In this work, we describe on the one hand how we developed an ontology of radiology procedures based on the RadLex Playbook and integrating the relevant classes of the RadLex ontology. On the other hand, we propose a proof of concept of the use of this procedures ontology by developing a tool to automatically classify imaging procedures from DICOM medical imaging metadata.

In this paper, we first describe the creation of our radiological procedures ontology and the development of our classification tool, explaining how we dealt with the challenges we faced. We then detail our results before discussing the limits of our design. Finally, we discuss the opportunities offered by this work, and particularly the ontological approach.

2. Material and methods

2.1. Creation of the ontology of radiological procedures

The Playbook is a CSV file defining more than 4400 procedures identified by a unique RadLex Playbook ID (RPID). Procedures are described through some text fields (some created manually, others automatically) and by a set of variables whose values are references to RadLex ontology elements. Concretely, the last column named "RIDS" contains a list of RadLex identifiers separated by a vertical bar which allows to link fields from MODALITY to VIEW_4 with RadLex entities. Figure 1 shows a few columns of the Playbook, in this example the value "RID13060" in the set of values of column RIDS refers to the MODALITY column and represents the "magnetic resonance imaging" class of the RadLex ontology.

RPID	SHORT_NAME	AUTOMATED_SHORT_NAME	MODALITY	BODY_REGION	MODALITY_MODIFIER	ANATOMIC_FOCUS	RIDS
RPID2599		XRAY LE 1-2VWS ANKLE BILAT	XR	LOWER EXTREMITY	1 - 2 VIEWS	ANKLE	RID10345 RID13060 0 RID2638 0 0 0
RPID2600	XR Ankle 1-2V	XRAY LE 1-2VWS ANKLE	XR	LOWER EXTREMITY	1 - 2 VIEWS	ANKLE	RID10345 RID13060 0 RID2638 0 0 0
RPID2601		XRAY GUIDE MAJ JNT ASP	XR		GUIDANCE	MAJOR JOINT	RID10345 RID13060 0 0 0 0 0 RID

Figure 1. Sample of the RadLex Playbook.

First, to avoid working on the whole RadLex ontology, we extracted the subset of classes that are used in the Playbook. In order to maintain the hierarchical organization, we also extracted all the parents of each of these classes. We also extracted their descendants, since we believed that the latter could be useful during the process of matching DICOM metadata with the ontology classes.

To integrate the procedures into the ontology, we naturally chose to create a new class for each radiological procedure of the Playbook identified by a RPID. Text fields then became annotation properties while fields containing RPIDs became object properties. In order to conform to the RadLex ontology practice, we defined a Preferred_name annotation property. We assigned it with the value of the column SHORT_NAME (if not filled in, we used AUTOMATED_SHORT_NAME). All text columns from the Playbook were then converted to annotation properties with their original name (see Fig 2). For "RID columns" (from MODALITY to VIEW_4), filled in fields were converted to "owl:someValuesFrom" restriction on the "subClass of" definition of the new class (compare Fig 1 and 2). Some columns have several numbered versions. For example, 5 columns are defined for the body region, from BODY_REGION to BODY_REGION_4 (as seen on Fig 1) because an imaging procedure could cover several body regions. We have not considered a hierarchy within these columns and defined a single object property for these cases. Therefore, each field BODY_REGION_X being filled-in resulted in the definition of a new restriction has_BODY_REGION on the newly created class. The only specific item is the POPULATION field for which we also declared an "owl:allValuesFrom" restriction, as the patient can't be in several categories at once.

Finally, we defined an "owl:equivalentClass" clause gathering all these restrictions and adding the fact that the class is a "procedure" (already defined in RadLex). The MODALITY_MODIFIER column sometimes contains information on the number of views used for the procedure. We thought a data property would be more meaningful and we defined the "nb_views" data property. So, the values of MODALITY_MODIFIER

describing the number of views were specifically converted to constraints using this data property. Figures 1 and 2 show how the value “1 – 2 views” for MODALITY_MODIFIER becomes constraints on nb_views in the “owl:equivalentClass” clause in the resulting ontology. All these newly created classes become subclasses of “playbook_procedure” defined as a subclass of procedure.

The screenshot displays the following information:

- Annotations:**
 - AUTOMATED_LONG_DESCRIPTION:** A radiography RADIOLOGY ORDERABLE 1 - 2 VIEWS procedure focused on the ANKLE in the LOWER EXTREMITY
 - AUTOMATED_SHORT_NAME:** XRAY LE 1-2VWS ANKLE
 - LONG_NAME:** XR Ankle 1-2 Views
 - Preferred_name:** XR Ankle 1-2V
 - RPID:** RPID2600
- Description:** 'XR Ankle 1-2V'
- Equivalent To:**
 - procedure and (has_ANATOMIC_FOCUS some ankle) and (has_BODY_REGION some 'lower extremity') and (has_MODALITY some 'projection radiography') and (nb_views some xsd:integer[>= 1, <= 2])
- SubClass Of:**
 - has_ANATOMIC_FOCUS some ankle
 - has_BODY_REGION some 'lower extremity'
 - has_MODALITY some 'projection radiography'
 - has_MODALITY_MODIFIER some '1-2 views'
 - playbook_procedure

Figure 2. Example of Playbook procedure integrated into our ontology.

As our use cases concern French data, we had to adapt our ontology by translating textual properties as well as possible.

The Unified Medical Language System (UMLS) is a meta thesaurus incorporating many biomedical terminologies. The elements of the various references are aligned and often benefit from translation efforts in many languages. The RSNA made the effort to align many classes from the RadLex ontology on the UMLS through the annotation property UMLS_ID which made our task easier. We used the freely available file MRCONSO.RRF containing all elements of all source vocabularies in the UMLS. A small extract of this file is provided in Fig 3, showing elements coming from different sources and translated in various languages. We tried to link RadLex classes to UMLS codes using the UMLS_ID property or to get an exact matching between the Preferred_name property and the UMLS label. For every matching code, we extracted all French labels described in the UMLS file and added them to the RadLex class using the Synonym_french annotation property that we created.

```

C0037303|ENG|S|L0037303|VC|S0425684|N|A0489194|||2715-5950|CSP|PT|2715-5950|skull|0|N|256|
C0037303|ENG|S|L0037303|VC|S1018639|Y|A23972546||C12789|NCI_CDISC|PT|SDTM-LOC|SKULL|0|N|256|
C0037303|ENG|S|L2545181|PF|S3005686|N|A2665209|||UWDA|PT|71325|Set of bones of cranium|0|N|256|
C0037303|FIN|P|L5680220|PF|S6513281|Y|A13411734||M0019947|D012886|MSHFIN|MH|D012886|Kallio|3|N||
C0037303|FRE|P|L3253461|PF|S3780993|Y|A7447501||M0019947|D012886|MSHFRE|MH|D012886|Crâne|3|N||
C0037303|GER|P|L3287242|PF|S3814835|Y|A7531090||M0019947|D012886|MSHGER|MH|D012886|Schädel|3|N||
C0037303|GER|S|L1243730|PF|S1485676|Y|A27687821||M0019947|D012886|MSHGER|ET|D012886|Cranium|3|N||

```

Figure 3. Sample from the UMLS.

2.2. *Ontology-based classification of radiological procedures*

The DICOM standard organizes medical imaging information using a 4-level hierarchy. The PATIENT level contains the demographic data of the patient. The STUDY level describes an examination, this is the level we map to the codes of the Playbook. For the same imaging examination (called DICOM Study in the DICOM jargon), the practitioner may use several modalities in many configurations, the studies are therefore usually composed of several SERIES, each describing the results of each imaging operation (e.g. acquisition, processing). Finally, these SERIES are composed of elements of the IMAGE level each describing one image. Figure 4 shows this hierarchical organization, the blue part designates a DICOM study, i.e. a radiological procedure. At each level, the data is described by a sequence of key/value elements (in which the value can itself be a sequence of such elements) where keys are DICOM attributes.

In each healthcare facility, image data is managed differently and the DICOM data elements are filled with varying degrees of accuracy. The “Study Description” data element is the one most used by the operators and the whole radiology system. It is generally filled in with local codes, specific to the institution. These codes are known by the operators and allow an automatic pre-filling of the DICOM attributes. The RadLex RPIDs have been designed to be mapped with these local codes and thus allow radiology system from several institutes to communicate. The ACR Dose Index Registry project is an example of this use of the Playbook [6]. The process of mapping local codes on the playbook is a manual task and some initiatives aim to assist this mapping in a semi-automatic approach [7].

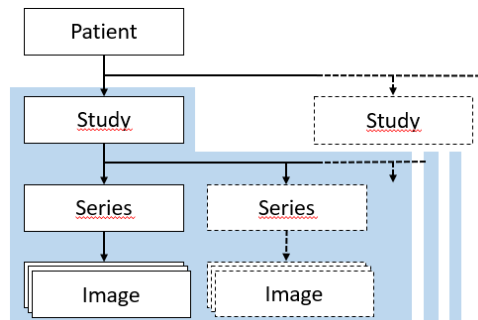


Figure 4. DICOM's hierarchical organization.

In this proof of concept, we worked from the raw data of each examination and tried to place it into the RadLex ontology. Our data-driven approach aims to enrich our data by getting the most out of each instance of imaging procedure.

We place ourselves at the moment of the extraction of DICOM data from the hospital PACS and before inserting data into our CDW. At this stage, each examination should result in an individual instantiating a class of radiological procedure from our ontology.

We first identified a list of DICOM attributes of interest. Among these, some - should- have normalized values, for instance “Image Laterality” has enumerated values: “R = right L = left U = unpaired B = both left and right”. Therefore we designed our classifier to manage each case by associating the right class, e.g. when the value “B” is

encountered in attribute “Image Laterality”, we add a restriction on property `has_LATERALITY` aiming at RadLex class “bilateral” (RID5771). Using the “Study Date” and “Patient’s Birth Date”, we similarly set a `has_POPULATION` restriction depending on the patient’s age (for instance, targeting class “neonatal” if the patient is less than a month old).

The other data elements are filled in using free text data. By using defined attributes whose values are descriptive (“Study Description”, “Series Description” and so on), we extracted all words and groups of 2 or 3 words. From these samples, we removed those that appear to be in a negative context, checking if they were preceded by words like “without”, “non”, “no”, etc. We then enriched this list using a mapping of acronyms, abbreviations, and codes to keywords. This mapping is a json file that we populated manually, for instance associating “TAP” to keywords “thorax”, “abdomen” and “pelvis”. The next step was to find RadLex classes matching the elements of this enriched list of samples. To do so, we created an inverted dictionary that allowed to find RIDs from the values of the annotation properties “Preferred_name”, “Acronym”, “Synonym” and “Synonym_french”.

Once we had a set of RIDs, we checked which object properties were likely to point to them. For example, RID10312 “MRI” was only targeted by object property `has_MODALITY`. For each of such property, we added an object property assertion aiming at a created individual of the target class (IRI built by concatenating the Preferred_name and a generated UUID). Figure 5 shows how a procedure resulted in an individual with 9 object property assertions.

The screenshot shows a web interface for an ontology individual. At the top, the URI is `1.2.250.1.74.20200407163000.1000089161203`. Below this, there are tabs for "Individual Annotations" and "Individual Usage". The "Annotations" section is currently empty. The "Description" is `1.2.250.1.74`. The "Property assertions" section is expanded, showing a list of 9 assertions:

- `has_ANATOMIC_FOCUS bone_organfdf3625f-c4c1-4da0-bae2-b8f61cfd8f7`
- `has_ANATOMIC_FOCUS lung61ea9957-5a41-4648-97a2-64fcdcae9d41`
- `has_TECHNIQUE sagittal_plane5ae6a647-50bc-4e4c-a52a-991841e44aad`
- `has_BODY_REGION bone_organb7ecea17-882c-473b-a925-3802dcec9839`
- `has_BODY_REGION spine722a74f6-de75-439f-b64e-8a6afd73376d`
- `has_BODY_REGION skullb56ea1a3-b697-4f3d-b166-d8c11bd2348c`
- `has_MODALITY computed_tomography5ae5b7f2-b1ac-4d5a-a551-ca5da75d7703`
- `has_ANATOMIC_FOCUS skull6c51764c-a05b-42f2-aa86-2d17a2d8e64d`
- `has_MODALITY_MODIFIER computed_tomographya15fab34-c1b7-44c8-92fa-52db7531de50`

On the left side, under "Types", the following classes are listed:

- `procedure` (highlighted in yellow)
- `'CT BONE'` (highlighted in yellow)
- `'CT SPINE'` (highlighted in yellow)

Below the types, there are options for "Same Individual As" and "Different Individuals", both currently empty.

Figure 5. An individual in the procedures ontology, created from a real imaging examination.

Once the instances available, we can simply deduce the classes to which they belong using an OWL reasoner, to this end we used HerMiT [8]. An example is provided in Fig 5, the individual has been considered as both a “CT BONE” procedure and a “CT SPINE” procedure, based on the assertions we produced with our extraction program. The development works presented in this part have been done with java and the Jena library.

3. Results

3.1. The procedures ontology

The ontology of the procedures we produced contains 16506 classes, 4402 classes resulting from the integration of the Playbook, and 12104 classes extracted from the RadLex ontology. We defined eleven new object properties and one data property (nb_views) to describe these classes. Fig 6 shows the mapping between the columns of the Playbook and our new object properties.

PlayBook Column	Playbook entries (RPID) using this column	Property assertion in "equivalent To" clause in the
MODALITY	4402	has_MODALITY
PLAYBOOK_TYPE	4402	N/A
POPULATION	88	has_POPULATION
BODY_REGION	3735	has_BODY_REGION
BODY_REGION_2	622	has_BODY_REGION
BODY_REGION_3	216	has_BODY_REGION
BODY_REGION_4	87	has_BODY_REGION
BODY_REGION_5	45	has_BODY_REGION
MODALITY_MODIFIER	2178	has_MODALITY_MODIFIER
MODALITY_MODIFIER_2	246	has_MODALITY_MODIFIER
MODALITY_MODIFIER_3	8	has_MODALITY_MODIFIER
PROCEDURE_MODIFIER	333	has_PROCEDURE_MODIFIER
PROCEDURE_MODIFIER_2	2	has_PROCEDURE_MODIFIER
ANATOMIC_FOCUS	2601	has_ANATOMIC_FOCUS
ANATOMIC_FOCUS_2	202	has_ANATOMIC_FOCUS
LATERALITY	923	has_LATERALITY
REASON_FOR_EXAM	1104	has_REASON_FOR_EXAM
REASON_FOR_EXAM_2	92	has_REASON_FOR_EXAM
REASON_FOR_EXAM_3	0	has_REASON_FOR_EXAM
TECHNIQUE	246	has_TECHNIQUE
PHARMACEUTICAL	1874	has_PHARMACEUTICAL
PHARMACEUTICAL_2	1	has_PHARMACEUTICAL
VIEW	142	constraint on nb_views
VIEW_2	65	constraint on nb_views
VIEW_3	24	constraint on nb_views
VIEW_4	8	constraint on nb_views

Figure 6. Mapping of the Playbook columns onto the object properties of the ontology.

4623 French synonyms have been found and 1301 classes have been translated with at least one French synonym. However, we sometimes detected errors due to the matching that is done on acronyms. Thus, MRI was associated with "Mauritanie" (Mauritania) or "arriération mentale" (Mental Retardation). Our ontology is freely available [here](#).

3.2. The proof of concept

We tested our program with data from two different institutions, each time by picking up the characteristics of all the radiological procedures that had been produced in one day on modalities "US", "CT" and "MR" (for ultrasound, computed tomography,

and magnetic resonance, respectively). Figure 7 shows how object properties were created on our radiological procedure individuals. For example, on the first institution, 4 individuals defined a `has_REASON_FOR_EXAM` property; these properties took 3 different values (“transplanted organ”, “pregnancy” and “injection”).

our Object Property	Institution 1		Institution 2	
	number of individuals using this Object Property	number of different Radlex classes targeted by this Object Property	number of individuals using this Object Property	number of different Radlex classes targeted by this Object Property
<code>has_MODALITY</code>	122	4	75	5
<code>has_MODALITY_MODIFIER</code>	58	2	63	2
<code>has_BODY_REGION</code>	41	6	57	9
<code>has_ANATOMIC_FOCUS</code>	29	11	41	5
<code>has_REASON_FOR_EXAM</code>	4	3	8	5
<code>has_POPULATION</code>	4	1	2	1
<code>has_TECHNIQUE</code>	0	0	26	1

Figure 7. Detail of the created procedure individuals.

On the first one we retrieved 122 instances; on average these instances are described by more than 2 object properties. On the other we retrieved 75 instances, described by 5 object properties on average. In both cases the execution time of the request was less than 30 seconds and the creation time of the instances from the raw data was about five seconds. The reasoner HermiT takes 20 seconds to load and classify our individuals.

Figure 5 shows an example of classification where the individual was classified in “CT Bone” and “CT Spine”. Of all the procedures received in both institutions, the program detected at least the modality. In the first institution, out of 122 individuals, 4 were classified as 2 different Playbook procedures, 92 were classified in 1 procedure and 26 were not classified. For those that were not classified, the only information collected was the procedure’s modality worth “US” (ultrasound) but there is no “US” procedure that would describe a procedure that we simply know has an “ultrasound” modality (as is the case for MR). In the second institution, out of 75 individuals, 42 were classified in one class, 30 have two classes and 3 have 4 classes.

The last result of this work is our java program. To sum it up, it allows us to perform 4 tasks: 1) From an ontology and a list of class identifiers, extract the subset of these classes with their direct ascendants and descendants to form an ontology extract; 2) Merge the Playbook and the RadLex extract to form a procedures ontology; 3) From an ontology and the UMLS file, enrich the ontology with French synonyms (configurable for another language); 4) Run a DICOM query on a PACS, parsing the data using a keyword mapping file and producing an RDF file of individuals that a classifier can use to classify in our procedures ontology.

4. Discussion

4.1. The procedures ontology

We believe that our radiological procedures ontology can be very useful and have demonstrated a use case for it. This work has been done with the idea of building on the extensive work done by many experts in the development of RadLex. However, while the advantages of RadLex are many, there are a few caveats. The description of anatomy

for example could be improved even if an effort was made to map some terms to the FMA [9]. More globally, it would be interesting to have an alignment of RadLex with formal ontologies such as BFO and other core ontologies from the Open Biological and Biomedical Ontologies Foundry (OBO) [10].

The richness of annotation properties “Preferred_name”, “Acronym”, “Synonym” and “Synonym_french” is important as we rely on it to match source data and RIDS, it however needs improving. For example, our program could not handle the word “épaule” (shoulder) because the Playbook does not use the RadLex class “shoulder” (RID39518) but “shoulder girdle”(RID1852) and the latter could not be translated.

Entries in the Playbook cover the different modalities differently. An “MR” entry makes it possible to assign a class to all MRI scans even if no other information is available, but such an entry does not exist for ultrasound. For CT, such an entry exists (RPID88) but its Preferred_name is “CT Guide needle place” which reveals that this procedure was intended to model CT Guidance for Needle Placement, although this motivation is not explicit in, e.g., the REASON_FOR_EXAM column of the Playbook.

The RadLex Playbook is not used in France, nor any other radiology procedures vocabulary, but inter-institutional grouping projects could gradually lead RadLex or other standard vocabulary to take hold.

As the Playbook is compiled manually and focuses more on classification than on description, the number of columns mapped remains limited (even if it is already very large). Also, it is sometimes difficult to have a clear definition of the role of a column, for example MODALITY_MODIFIER which points at classes of many different types. Similarly, in [2], REASON_FOR_EXAM is defined as: “Information about the clinical indication, patient diagnosis, clinical status (e.g., postoperative), an intended measurement, altered anatomy (e.g., endograft), or some other purpose of the study (eg, screening)”. There is obviously significant overlap between columns MODALITY_MODIFIER_X, REASON_FOR_EXAM_X and TECHNIQUE, possibly also resulting from the merging work with LOINC.

Our work being based on ontological reasoning for classification and on annotation properties to make the link between keywords and classes, it should rely on axioms definitions that are fully consistent, i.e. non-overlapping and where the use of subsumption is perfectly correct. An alignment with BFO, the use of methods based on OBO principles for the alignment with other ontologies and a restructuring work on the playbook could therefore be a way to improve our system and could be done as a continuation of this work.

A quick look at the results already showed the value of our approach. For example, an examination in which we found the term “mammo” was associated with the RID “mammography”. The latter is a sub-class of RadLex “radiography projection” (RID10345) which is a potential target class of the object property has_MODALITY, the newly created procedure individual has therefore been linked to a mammography individual via has_MODALITY. The reasoner classified the examination with the procedure code “X-RAY” (RPID2501) because it had a radiography projection as its modality.

By associating an institution's local codes with the RPIDs like it is done in the ACRDose project, one could situate the labeled exams within RadLex and thus take advantage of the power of this ontology. However, our approach is to take advantage of the power of the ontology for classification itself since we place the concrete objects in RadLex, which then allows classification. Thus, if RadLex changes, we simply need to regenerate our ontology.

4.2. *The proof of concept*

Further validation of this work will be required before it can be used in practice in the CDW, where its usefulness will be further evaluated. If the “ontologization” of the Playbook is rather successful, the use of the DICOM data raises challenges that are not completely solved by our solution and will still require work on the development of our algorithm and the parsing of the input data.

More information could be drawn from the DICOM metadata. DICOM attributes are not very filled in, it is mainly the description fields that enable classifying the procedure. Moreover, there is a difference of method between the two institutions: in one of them the DICOM data element “Body Part Examined” is filled in (with free text instead of DICOM enumerated values), in the other it is a simple copy of the free text data element “Study Description”, which is problematic because this data element is 16 characters long, so the value is often truncated. Beyond the management of acronyms, abbreviations and codes, we faced the problem of adjectives. The difference in classification between institutions depends mainly on the fact that the first one uses adjectives, for example "encéphalique" that does not match in our ontology. A possible solution would be to manage this case like the others by adding in our mapping entries such as "encéphalique" to the keyword "crâne" (skull), or to find an already existing resource providing this link.

One of the main problems is the management of classes that describe the absence of a thing or an action. To assert that an exam is “Without IV contrast”, the data should be analyzed in depth (perhaps down to the IMAGE level). If a local code defines this “lack of IV contrast”, we could associate it in our keyword mapping file to the label “imaging without IV contrast” so that the algorithm would correctly add an assertion on has_MODALITY_MODIFIER to RID28768 (“imaging without IV contrast”).

One of the advantages of our solution is that the refining axes are reduced to the constitution of configuration files. There is still mapping work on local codes, acronyms, abbreviations and adjectives, but this work is very generic and does not depend on a specific ontology or institution. In the long term, one could imagine replacing this mapping with a Natural Language Processing approach. It was also found that the use of DICOM could be optimized in general and is not the same everywhere. In practice most of the information is in the Study Description data element. Our approach would be sensitive to an improvement in the filling and quality of the source data.

By improving and validating the accuracy of our program we can envisage other uses. If our program is unable to classify a radiological procedure or if it classifies it in several very different classes, it may be that it is a new examination not covered by the Playbook yet, or that the DICOM file is filled out incorrectly. With improved accuracy, it could be used to help detecting new uses or errors.

5. Conclusion

Starting from the problem of classification and interoperability in our CDW, we have designed the first ontology of the RadLex Playbook in OWL. By merging the Playbook with the relevant extract of the RadLex ontology, we were able to capitalize on the great work already done by the actors of these projects. We have given an example of how we have enriched our ontology resource, translating it in another language via UMLS.

Finally, we created a concrete use case of our ontology of procedures that we were able to test with data from two different institutions.

There are many ways to improve our ontology, its translation and our proof of concept, but the data-driven approach is already giving interesting results. The use of such a system offers new opportunities within the framework of CDW given the importance of interoperability and the increasing place of imaging.

Acknowledgement

We would like to thank the French National Research Agency (ANR), for funding this work inside the LabCom LITIS (Laboratoire d'Interopérabilité, de Traitement et d'Intégration des données de Santé) project (grant no. ANR-17-LCV1-0004).

References

- [1] Madec J, Bouzillé G, Riou C, et al. eHOP Clinical Data Warehouse: From a Prototype to the Creation of an Inter-Regional Clinical Data Centers Network. *Stud Health Technol Inform.* 2019;264:1536 - 1537. doi:10.3233/SHTI190522
- [2] Wang KC, Patel JB, Vyas B, Toland M, Collins B, Vreeman DJ, Abhyankar S, Siegel EL, Rubin DL, Langlotz CP. Use of Radiology Procedure Codes in Health Care: The Need for Standardization and Structure. *Radiographics.* 2017 Jul-Aug;37(4):1099-1110. doi: 10.1148/rg.2017160188. PMID: 28696857; PMCID: PMC5548452.
- [3] Wang KC. Standard Lexicons, Coding Systems and Ontologies for Interoperability and Semantic Computation in Imaging. *J Digit Imaging.* 2018 Jun;31(3):353-360. doi: 10.1007/s10278-018-0069-8. PMID: 29725962; PMCID: PMC5959830.
- [4] Vreeman DJ, Abhyankar S, Wang KC, Carr C, Collins B, Rubin DL, Langlotz CP. The LOINC RSNA radiology playbook - a unified terminology for radiology procedures. *J Am Med Inform Assoc.* 2018 Jul 1;25(7):885-893. doi: 10.1093/jamia/ocy053. PMID: 29850823; PMCID: PMC6016707.
- [5] NEMA PS3 / ISO 12052, Digital Imaging and Communications in Medicine (DICOM) Standard, National Electrical Manufacturers Association, Rosslyn, VA, USA (available free at <http://medical.nema.org/>)
- [6] Morin RL, Coombs LP, Chatfield MB. ACR Dose Index Registry. *J Am Coll Radiol.* 2011;8(4):288 - 291. doi:10.1016/j.jacr.2010.12.022
- [7] Mabotuwana T, Lee MC, Cohen-Solal EV, Chang P. Mapping institution-specific study descriptions to RadLex Playbook entries. *J Digit Imaging.* 2014 Jun;27(3):321-30. doi: 10.1007/s10278-013-9663-y. PMID: 24425187; PMCID: PMC4026460.
- [8] Glimm B, Horrocks I, Motik B, Stoilos G., Wang Z. (2014). HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning*, 53, 245-269.
- [9] Mejino JL, Rubin DL, Brinkley JF. FMA-RadLex: An application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology. *AMIA Annu Symp Proc.* 2008;2008:465 - 469. Published 2008 Nov 6.
- [10] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ; 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 Nov;25(11):1251-5. doi: 10.1038/nbt1346. PMID: 17989687; PMCID: PMC28140