Early steps of an Ontology for Magnetic Resonance Imaging: MRIO

Lucas M. Serra [1], Michael G. Dwyer [2], William D. Duncan [1], Alexander D. Diehl [1]

Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, USA[1]; Buffalo Neuroimaging Analysis Center, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, USA[2].

Abstract

The Magnetic Resonance Imaging Ontology (MRIO) is an application ontology that represents numerous entities in the domain of magnetic resonance imaging (MRI) including MRI analysis and MRI sequences. Data from clinical trials MRI protocols were used to create the axioms of these MRI sequences. We have also created means for automatically loading MRI headers as new ontology instances and demonstrate the ability to query data in MRIO. The current work represents the beginnings of a full-fledged imaging ontology and automated analysis pipeline, which we plan to further develop. Future iterations of the project will include a stream-lined user-interface for querying and improved capability in classifying image types.

Keywords:

MRI ontology; imaging informatics; MRIO.

Introduction

The Fundamentals of Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a mainstay of modern medicine that has rapidly integrated itself into a myriad of diagnostic algorithms and has proven itself as a valuable component of healthcare due to its versatility and accuracy. However, these features come at the cost of price and complexity. MRI is a nuanced technology and, when approaching methods for representing its components in an ontology, merits an understanding of the fundamental principles of magnetic resonance. MRI is based upon the same physical principles that underlie nuclear magnetic resonance and is predicated upon on the notion of "spin". Spin gives particles, like protons, their angular momentum and a magnetic moment (1). Protons therefore have magnetic fields which align with applied external magnetic fields. By interrogating these proton spins with radiofrequency pulses and recording the responses, MRI is able to infer many different properties of the underlying tissue.

The Anatomy of an MRI Machine

Modern MRI machines are composed of a primary superconducting magnet that supplies the main magnetic field, a gradient coil to alter the primary magnet's field and encode spatial information, and a set of radiofrequency (RF) coils to create pulses and receive signals. If we consider an analogy where the protons are the needle of a compass, the RF coil's function is somewhat similar to nudging the needle with a finger and timing how long it takes for the needle to re-right itself. As the protons re-align themselves with the applied magnetic field, they release energy. Protons can release energy to their surroundings, which is referred to as spin-lattice relaxation or T1 relaxation. Alternatively, protons can become out of phase with each other. This is called spin-spin relaxation or T2 relaxation. Depending on which of these effects dominates an image determines whether we designate an image as a "T1 image" or a "T2 image". The aforementioned effects alter the net magnetic vector within the machine, which is captured as electrical impulses by the RF coil. In addition to these "classical" image contrasts, the field of MRI physics has discovered many other sources of tissue contrast that can be elucidated by variations in the standard pulse sequence regime. Together, these various contrasts enable fine discrimination of tissue composition that is not possible with other imaging modalities, which has cemented MRI as the premiere imaging option for pathologies affecting soft tissues.

Expanding Use and Standards

MRI is an often-used component in a physician's toolkit especially in the US which boasts the second highest number of MRI machines per capita globally (2). MRI has broad clinical and research applications ranging from traumatic brain injuries to osteoarthritis to malignancy. Within the past two decades, the use of imaging across healthcare has risen dramatically, and has been partly fueled by physicians who purchase MRI machines for their practices and consequently order more scans (3, 4). The growing use of imaging data has necessitated improvements in imaging standards and protocols. Healthcare professionals and researchers working within the field of imaging wisely adopted a standard file format for medical images decades ago. Digital Imaging and Communications in Medicine (DICOM) is used worldwide to store and transmit medical images. (5). In order to further augment the standardization and interoperability introduced by DICOM, centers involved in clinical trials often adopt detailed protocols, which state specific image parameters and tolerances for use during data collection. These data exist as numbers and text in the metadata fields of a DICOM header.

Problems Facing the Field

I. With increased use and widening adoption comes ever-growing volumes of data that must be catalogued, managed, and analyzed. Despite the progress made in standardizing medical images, there exist numerous challenges in the management of imaging data, which the use of an ontology helps to mitigate. The metadata fields of a DICOM file frequently represent non-explicit knowledge using ambiguous language. For instance, one of the fields in the DICOM header is labeled 'PulseTime'. The preceding fields deal with cardiac aspects of the scan such as ‘CardiacRepetitionTime’ and ‘ImagesPerCardiacCycle’ which may lead one to believe that ‘PulseTime’ relates to the pulse or heart rate of the patient. This is complicated by later fields that reference RF pulses but instead do so using language like ‘PulseSequence’. This makes it challenging for a user who is unfamiliar with the domain to use the data. Fully understanding the intended meaning of the data fields involves deep knowledge of the latest version of the DICOM specifications, use of a third-party website, or consultation with a domain expert. Among the most important issues is a lack of consensus about the exact parameters that make up a specific image type, which is partly confounded by intermachine and inter-operator variability. The Alzheimer's Disease Neuroimaging Initiative (ADNI) maintains highly detailed MRI scanner protocols for use in its clinical trials and illustrates this variability well (6). The ADNI 3 protocols define the MRI acquisition parameters for capturing a sagittal 3D fluid-attenuated inversion recovery image of a human brain in several different machines from different vendors. In a General Electric 25 MRI machine, the echo time (TE) is 119.0ms, the repetition time (TR) is 4800.0ms and the inversion time (TI) is 1451ms while in a Siemens Magnetom Verio machine the parameters are 442ms, 4800ms, and 1650ms for TE, TR, and TI respectively. Although both machines are 3 tesla MRI machines and attempting to capture the same image, their TE parameters are quite different. Moreover, even small changes in these parameters can result in radically different images and associated image types. Broadly speaking, we currently do not have effective methods for transitioning from these elementary imaging parameters to higher semantic levels. If we borrow an analogy from biology, these imaging parameters are similar to the nucleotides of DNA where different sequences can code for the same codons and proteins. As of yet, we lack an elegant way to determine these proteins or their functions from their constituents. These factors can result in problems with interoperability when combining large sets of MRI images and the requirement to write complex and cumbersome queries to create retrospective cohorts.

Imaging Ontologies

The current work is not the first ontology in the domain of MRI images, and a handful of past studies have created MRI-related ontologies. NeuroLOG or OntoNeuroLOG is a French multi-level ontology created to integrate neurological resources from multiple academic centers and uses DOLCE as its upper level ontology (7). NeuroLOG covers a wide array of brain-centric investigation-related entities including MRI (8). A more recent MRI ontology covered MRI simulations and modeled the fundamental processes of the RF pulses that form sequences (9). Lastly, the DICOM controlled terminology is available on BioPortal and consists of every term used in the DICOM file format along with their definitions (10). These works suffer from limitations in accessibility and usability. NeuroLOG is inaccessible through the paper’s provided links and what is viewable through snapshots of the ontology show missing textual and logical definitions for represented entities. NeuroLOG also uses DOLCE, which restricts its interoperability with the multitude of existing OBO Foundry ontologies that are grounded in the Basic Formal Ontology. Interoperability with OBO Foundry ontologies is an important feature that promotes reuse and prevents the creation of isolated “data siloes”. The ontology covering MRI simulations and sequences did not publish their ontology in any form. The DICOM controlled terminology, although published alongside ontologies on BioPortal, has a completely flat structure and some of its definitions are not crafted in the style preferred by the OBO Foundry. Additionally, all these ontologies seem to not cover the higher levels of abstraction that we desire in our ontology.

In the current work, we have developed the MRI ontology (MRIO) to represent MRI analyses, sequences, images, and machines using metadata from DICOM files to create axioms. We have also created methods for extracting this information from DICOM headers and automatically creating new ontology instances.

Methods

Ontology Construction

MRIO was created with the latest version of Protégé (5.5.0) (11). The HermiT (1.4.3.456) reasoner plugin was used for inference (12). Our ontology was built with certain principles in mind, such as resource identifiers, textual definitions, and openness, all of which are outlined by the Open Biological and Biomedical Ontology (OBO) Foundry (13). Following these principles, MRIO uses BFO as its upper level ontology and re-uses existing ontologies like the Ontology for Biomedical Investigations (OBI) and the Information Artifact Ontology (14-16). MRIO adds 70 new terms, most with well-constructed textual and logical definitions to represent multiple aspects of MRI images. Around two dozen terms were reused from OBI and IAO as upper level terms or in relations. Our ontology was constructed in both a top-down and a bottom-up approach. The entities we deemed most important in representing MRIs in an ontology are: the MRI image objective and the MRI sequences, followed by the MRI machine, the patient/evaluant, the MRI assay, and the MRI image itself. We consulted with domain experts in order to create the MRI analysis hierarchy. The most salient metadata on DICOM image files are “parameter specifications” or “acquisition parameters”, which describe RF pulse sequences. These parameters, implemented as data properties, were used in creating the axioms and computer-readable definitions of sequences. A GitHub repository containing the latest version of the ontology is available at: https://github.com/LucasSerra1/MRIO.git

Data Extraction

The scripts used in parsing MRI headers and MRI protocol files were written using the Python programming language. In essence, the scripts extract information from the DICOM headers and transform the information into instances of MRIO classes and relations. DICOM header data fields are first transformed into a spreadsheet. These fields are mapped to MRIO data properties. Numeric values are then read and associated with these data properties. The RDFLib (4.2.2) Python library was used to facilitate this transformation and automatically add graph nodes and new instances to our ontology from these mapped classes. To create the axioms that underlie the sequence types (Figure 2), a separate script was created that extracts parameter specifications from JSON files representing years of MRI study protocols used in clinical trials conducted at the Buffalo Neuroimaging Analysis Center. As no exact definitions for consensus sequence parameters exist in the DICOM specifications or in published literature, simple ranges were used to define sequence parameters and provide a survey of the data available. Minimum and maximum values were extracted across hundreds of entries to create the ranges that constitute the axioms of our sequence classes.

Our final output of the data extraction process was an OWL file containing 4 instances (representing a single DICOM header), 70 MRIO-specific classes, and 8 new data properties. The original data consisted of 300 text files containing 1000 entries for MRI protocols (17). This was distilled into 5 MRI pulse sequence classes in the final ontology. After modification with RDFLib, the ontology was loaded as a triplestore into the free version of GraphDB (8.9) Using SPARQL, we queried the data looking for images by their parameters (18, 19).

Results

The ensemble of these moving pieces is a pipeline that automatically loads DICOM headers and inserts them into a queryable MRI ontology created from a combination of domain expertise and parameter data extracted from clinical trial protocols. Figure 3 provides an overview of the gross structure of the ontology. As MRIO is built upon the foundations of OBI, it takes a similar approach in establishing relationships between the overall imaging process and the participants. Terms derived from OBI are in ovals while MRIO terms are in boxes. More specifically, 'magnetic resonance imaging pulse sequence' is define as a type of 'processed material' and stands in a 'part of' relation to the 'magnetic resonance imaging radiofrequency coil', which is an OBI 'measurement device'. As shown in Figure 3, both the MRI machine and a 'material entity' with the 'magnetic resonance imaging evaluant role' are the specified inputs of a 'magnetic resonance imaging assay'. This 'magnetic resonance imaging assay' term resides under the 'planned process' class and has 'magnetic resonance imaging datum' as the specified output. This data undergoes a 'magnetic resonance imaging data transformation', which in the real-world partly takes the form of a Fourier transformation and results in the final 'magnetic resonance imaging image'. The image is tied back to the sequence used and the subject of the scan using 'is about' relations.

Figure 1 depicts the structure of the MRI pulse sequences. Several new data properties were needed to fully represent sequence parameters: 'has TR', 'has TE', 'has inversion time', 'has flip angle', and 'has echo train length'. These entities were derived from BNAC MRI protocol specifications and represent settings configured on an MRI machine for the creation of an MRI image.

Figure 2 illustrates the type of query one is able to use with MRIO. With SPARQL, an investigator is able to hone in on well-crafted cohorts via sequence parameters as in this example or via a number of other axes.

Discussion

Our work contributes to imaging informatics in a number of ways. The automatic creation of ontology instances mitigates the laborious task of data entry. Our system also enables precise selection of cohorts from datasets of DICOMs and facilitates discovery of potential subgroups within imaging data. MRIO provides a structured semantic representation of many of the metadata fields found in the DICOM format. To this end, MRIO improves the interpretability of data field definitions without the need for external resources and elucidates some of the implicit knowledge found within this domain. MRIO's adherence to OBO Foundry principles also enhances interoperability with other similarly structured ontologies.

Despite these benefits, MRIO and its extraneous systems are currently limited in some respects. At present, MRIO can only process single DICOM headers, which must be loaded as text files. Furthermore, once new MRI instances are loaded, the HermiT reasoning engine in Protégé takes minutes to sort individuals and infer relations. This occurs with only a handful of DICOMs loaded. We are investigating methods to speed up the reasoning so we can scale the ontology appropriately. Our ontology also only captures a small selection of the vast number of data fields found within the DICOM file standard. We would also like to more fully develop the definitions of our classes. As a final limitation, our system requires that users understand SPARQL to write their queries and extract information from data loaded in triplestores, although our long-term plans include creating a web interface to simplify querying.

MRIO represents the beginnings of a full-fledged imaging ontology and automated analysis pipeline. There are many possibilities for future work and expanding the functionality of MRIO. With thousands of MRIs loaded from disparate data sets and institutions, it would be possible to better grasp which are the exact elements that make a "T1 image". This could occur either through community consensus or MRIO could provide high-quality data for machine learning or statistical treatments of this question. In later versions of our work, the query system could be improved with a natural language processing-based query system and a more stream-lined user interface that would obviate the need for users to know SPARQL.

Conclusion

MRIO is the only MRI ontology under active development. At present, MRIO enjoys a number of useful features and these initial steps provide a proof-of-concept for a much larger analytic platform with numerous uses.

ACKNOWLEDGMENT

AD was supported by 5UL1TR001412 (NCATS).
MD has received consultant fees from Claret Medical and EMD Serono, and research grant support from Novartis and Celgene.

Fig. 1. Sample MRI pulse sequence class

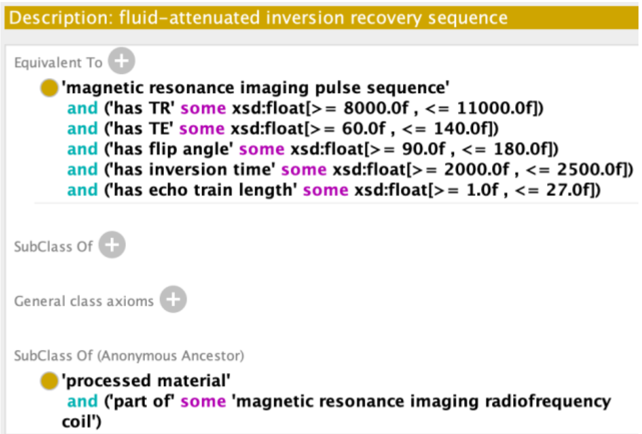
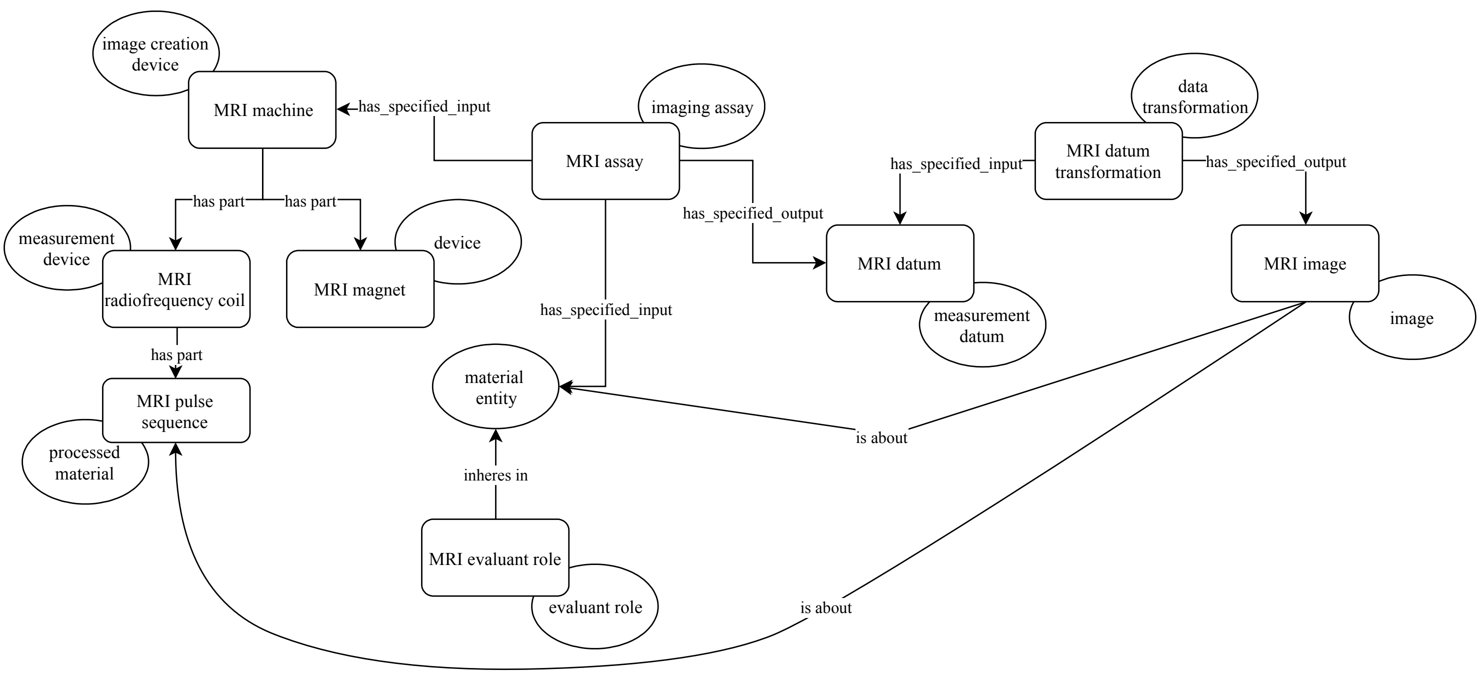


Fig 2. Example SPARQL query

```
select ?image ?TR ?TE ?flip
where {
    ?image <http://purl.obolibrary.org/obo/MRIO_0000376> ?TR
    Filter (?TR>=9000)
    ?image <http://purl.obolibrary.org/obo/MRIO_0000377> ?TE
    Filter (?TE<=140)
     ?image <http://purl.obolibrary.org/obo/MRIO_0000384> ?flip
    Filter (?flip=90)
}
```

Fig. 3. Gross and relational structure of MRIO



REFERENCES

1. Plewes DB, Kucharczyk W. Physics of MRI: A primer. Journal of Magnetic Resonance Imaging. 2012;35(5):1038-54.
2. Health expenditure indicators [Internet]. 2014. Available from: https://www.oecd-ilibrary.org/content/data/data-00349-en.
3. Agarwal R, Bergey M, Sonnad S, Butowsky H, Bhargavan M, Bleshman MH. Inpatient CT and MRI utilization: trends in the academic hospital setting. Journal of the American College of Radiology : JACR. 2010;7(12):949-55.
4. Baker LC. Acquisition of MRI equipment by doctors drives up imaging use and spending. Health affairs (Project Hope). 2010;29(12):2252-9.
5. Mildenberger P, Eichelberg M, Martin E. Introduction to the DICOM standard. European Radiology. 2002;12(4):920-7.
6. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology. 2010;74(3):201-9.
7. Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A. WonderWeb deliverable D18 ontology library. 2003.
8. Michel F, Gaignard A, Ahmad F, Barillot C, Batrancourt B, Dojat M, et al. Grid-wide neuroimaging data federation in the context of the NeuroLOG project. Studies in health technology and informatics. 2010;159:112-23.
9. Lasbleiz J, Saint-Jalmes H, Duvauferrier R, Burgun A. Creating a magnetic resonance imaging ontology. Studies in health technology and informatics. 2011;169:784-8.
10. Salvadores M, Alexander PR, Musen MA, Noy NF. BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. Semantic web. 2013;4(3):277-84.
11. Noy NF, Crubezy M, Fergerson RW, Knublauch H, Tu SW, Vendetti J, et al. Protégé-2000: an open-source ontology-development and knowledge-acquisition environment. AMIA Annual Symposium proceedings AMIA Symposium. 2003;2003:953-.
12. Glimm B, Horrocks I, Motik B, Stoilos G, Wang Z. HermiT: An OWL 2 Reasoner. Journal of Automated Reasoning. 2014;53(3):245-69.
13. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology. 2007;25(11):1251-5.
14. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The Ontology for Biomedical Investigations. PLOS ONE. 2016;11(4):e0154556.
15. Smith B, Ceusters W. Aboutness: Towards Foundations for the Information Artifact Ontology2015.
16. Arp R, Smith B, Spear AD. Building Ontologies with Basic Formal Ontology: The MIT Press; 2015. 248 p.
17. Antoniou G, van Harmelen F. Web Ontology Language: OWL. In: Staab S, Studer R, editors. Handbook on Ontologies. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004. p. 67-92.
18. Gueting RH. GraphDB: A Data Model and Query Language for Graphs in Databases. Informatik-Bericht 155. 1994.
19. Prud’hommeaux E, Seaborne, A. SPARQL Query Language for RDF. W3C Recommendation. January 2008.