Development of an Image Data Set Class: Its Role in Biomedical Imaging and Neuroimaging Research

Alexander Bartnik[1,2], Mackenzie Smith[2,3], Lucas Serra[2,3], William D. Duncan[4], Lauren Wishnie[3], Alan Ruttenberg[3], Michael G. Dwyer[1,2] and Alexander D. Diehl[3]

[1] *Buffalo Neuroimaging Analysis Center, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, USA*
[2] *Department of Neurology, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, USA*
[3] *Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, USA*
[4] *University of Florida College of Dentistry, Gainesville, FL, USA*

Abstract
Biomedical imaging is a widely used tool both clinically and for research. Though a standard digital format has existed in the biomedical imaging world across modalities for decades in the form of the DICOM specification, a formal representation of the kinds of data present in a biomedical image (acquired from CT, PET, MRI, etc.) is notably absent from biomedical ontologies, and annotation of biomedical imaging data is hindered by decentralization. This has contributed to the creation of large and unsorted biomedical imaging silos, preventing clinical and translational researchers from effectively sharing and analyzing their data. We present here the 'image data set' class, along with the 'image data set analysis' class, which we have developed to capture the processes of acquisition, annotation, and analysis of biomedical imaging data in an effort to better harness otherwise-latent imaging datasets. The 'image data set' class and several of its children are being contributed to OBI and originate from MRIO, an application ontology used to guide a neuroinformatics platform working to automate analysis of large MRI datasets and facilitate the translation of neuroimaging research into clinical science.

Keywords [1]
Image data set, data set, biomedical imaging, MRI, MRIO, OBI

1. Introduction

Digital biomedical imaging, a technique enabling physicians and researchers to visualize a subject's internal structural and functional anatomy, has been a mainstay of modern medicine for decades [1]. Techniques for imaging comprise several distinct modalities, including computed tomography (CT), positron emission tomography (PET), X-ray, as well as nuclear magnetic resonance (NMR) spectroscopy and magnetic resonance imaging (MRI). The number of biomedical imaging scans performed clinically increases year-over-year [2], adding to an ever growing mountain of imaging data and firmly establishing the role of biomedical imaging in the future of medicine and clinical research.

Biomedical imaging in the realm of clinical practice is noticeably different from its use in

ICBO 2022, September 25-28, 2022, Ann Arbor, MI, USA
EMAIL: adbartni@buffalo.edu (A. 1); mtsmith6@buffalo.edu (A. 2); lucasser@buffalo.edu (A. 3); wdduncan@gmail.com (A. 4); lmwishni@buffalo.edu (A. 5); alanruttenberg@gmail.com (A. 6); mgdwyer@buffalo.edu (A. 7); addiehl@buffalo.edu (A. 68
ORCID: 0000-0001-9676-7377 (A. 1); 0000-0002-9821-4132 (A. 2); 0000-0002-2104-0568 (A.3); 0000-0001-9625-1899 (A. 4); 0000-0002-7245-3450 (A. 5); 0000-0002-1604-3078 (A. 6); 0000-0003-4684-4658 (A. 7); 0000-0001-9990-8331 (A. 8)

research, with clinical images trading quality for general usability for diagnosis in an individual, while research often requires higher quality scans at a higher price point. Because imaging is expensive [3], its potential use in research is limited, while clinics produce many images with the help of reimbursement from insurance. As such, there has been a push in recent years to mobilize the large quantity of images acquired via clinical routine for use in research [4–9]. However, this data is often dispersed and left unsorted in data silos, and harmonizing all of this data would take considerable time and effort. There exists then a need for a standardized method of automatically sorting and annotating these large amounts of imaging data, which would allow researchers to more easily analyze and share data.

1.1. The DICOM Standard

The widespread use of biomedical imaging is due largely to the adoption of the Digital Imaging and Communication in Medicine (DICOM) standard [10]. Since its major release in 1993, DICOM has been the International Organization for Standardization (ISO) recognized digital format for biomedical imaging, allowing interoperability across modalities, scanner manufacturers, and healthcare systems. As such, DICOM governs the reporting of image acquisition, transfer, and display. Furthermore, DICOM encodes data pertaining to image acquisition (e.g. date, time, patient age, acquisition parameters, modality, etc.) in the file headers, similar to a JPEG but specifically tailored to biomedicine. There exist large datasets of biomedical images with the potential to be harmonized because they already have DICOM annotations. Additionally, the near universal use of DICOM has led to the creation of a large ecosystem of software and libraries for programmatically working with DICOM images and headers (e.g. pydicom [11] for python, Java DICOM Toolkit for Java [12], Image Toolkit (ITK) [13], etc.).

1.2. The Role for Ontology in Biomedical Imaging Datasets

The adoption of the DICOM standard and increasing clinical use of imaging has generated a huge amount of data, presenting unique opportunities to clinical and translational researchers. However, while DICOM offers standardized methods for reporting, storing, and transferring scans in the form of standard header tags, much of the data encoded in DICOM headers are semantic strings that often differ between institutions and manufacturers. As such, important information for sorting and analyzing data (e.g. scan type, series description, date of birth, etc.) are often disparate from one data set to the next, and different methods must be used to work with different data sets. Moreover, there is a high barrier of entry for working with imaging data, requiring experience in fields like computer science and informatics that clinical researchers might not have. The field of neuroimaging in particular has recently been moving towards the adoption of the Brain Imaging Data Set (BIDS) [14] specification, which is a prescribed format for naming and laying out directories of neuroimaging data. While newer than DICOM and focused on MRI, BIDS also targets additional modalities such as PET and provides a framework for analysis of BIDS datasets called "BIDS-apps" with little-to-no input required from the user. However, BIDS is practical rather than ontological, and its recommended practices still require experience in the fields of computer science and neuroimaging analysis.

Ontology poses an elegant solution to these problems, but no term in any OBO Foundry ontology quite encapsulates the kind of data found in biomedical imaging. The obvious term to use would be 'image' (IAO:0000101) from the Information Artifact Ontology [15], which is defined as "[…] *an affine projection to a two dimensional surface, of measurements of some quality of an entity or entities repeated at regular intervals across a spatial range, where the measurements are represented as color and luminosity on the projected on surface*." Analysis of a biomedical image requires data such as scanning modality (i.e. MRI vs. PET) and scan type (anatomical vs. functional) to be known beforehand, which may be encoded in a DICOM header just as an affine projection of pixels would be. Additionally, biomedical images are typically comprised of voxels, similar to pixels but representing data in three or more dimensions. Other candidate terms include 'image' (NCIT:C48179) and 'medical image' (NCIT:C19477) from the National Cancer Institute Thesaurus [16], but these terms lack useful axioms for working with or sharing imaging data.

1.3. An Ontological Representation of Imaging Data

In light of this, we developed the 'image data set' class out of a need for a more general term for an information content entity that more fully represents the different kinds of information found in biomedical imaging data. Additionally, we developed an 'image data set analysis' term to represent a standard process for analyzing and deriving data from biomedical images. Both of these terms originate from our work on the MRI Acquisition and Analysis Ontology (MRIO) [17], an application ontology developed to capture the neuroimaging research process primarily focused on MRI data (https://github.com/Buffalo-Ontology-Group/MRI_Ontology). However, because 'image data set' is a generic child of 'data set' (IAO:0000100), there is certainly potential for the class to be used with other biomedical imaging modalities and any other digital image format. The 'image data set' class, several of its subclasses specific to MRI, and the 'image data set analysis' class are in the process of being contributed to the Ontology for Biomedical Investigations (OBI) [18].

2. Development

The 'image data set' and 'image data set analysis' classes were developed as high-level terms to contain the different kinds of MR images and analyses present in MRIO. These lower-level terms were added as children to the 'image data set' class. The terms were developed in Protégé v5.5.0's [19] ontology editor, and automated reasoning was performed using the HermiT reasoner v1.4.3.456 [20]. Table 1 provides the higher-level terms and definitions presented here.

2.1. Image Data Set

We define 'image data set' as, "*A data set that is comprised of structured measurements about some entity and its associated metadata using pixels (2D), voxels (3D), or an arbitrary number of dimensions. An image data set can be the source from which an image is produced*." The intent of this class is to provide a general term that may be extended to all types of biomedical imaging data, since the two-dimensional definition of 'image' from IAO does not allow for the three- or four-dimensional data typically found in biomedical imaging. Additionally, biomedical images in the form of DICOMs contain additional data pertaining to the acquisition of the scan that are necessary for analysis. The 'image data set' class allows for the inclusion of these (meta)data in addition to what may be thought of as simply the image that the DICOM encodes.

We also developed a dichotomy under the 'image data set' class to represent raw data produced by the scanner as well as data that has been transformed in some way (i.e. into DICOM or as part of any other analysis), using the 'raw image data set' and 'computed image data set' classes, respectively. The lower-level MRI scan types from MRIO were added under a child class of 'computed image data set' called 'reconstructed magnetic image data set.' Each of these MRI data sets are the output of a 'magnetic resonance imaging assay' (OBI:0002985), which we have expanded to encode information pertaining to acquisition parameters for each scan type as found in DICOM headers. This provides us a higher-level 'image data set' class that can be used for any biomedical imaging modality as well as lower-level terms for different MRI scan types logically defined by standardized acquisition parameters.

In addition to the MRI specific 'image data set' terms, we developed terms for annotating brain region atlases and image segmentation, called 'brain region atlas image data set' and 'image segmentation map,' respectively.

Figure 1 provides an overview of the acquisition of a 'magnetic resonance image data set.'

2.2. Image Data Set Analysis

We also developed a class for formally representing the analysis of 'image data sets' called 'image data set analysis,' which we define as, "*The process of deriving a data item from an image data set using computer algorithms. The produced data item can be an image data set, data measurement, or any other data item.*" Different imaging analysis software and tools can be and have been added as subclasses to the 'image data set analysis' class, along with logical axioms that define the inputs and outputs of the analysis. For example, the 'FreeSurfer analysis' term (MRIO:0000515) has been added as a subclass to 'MR image segmentation analysis' (MRIO:0000662) to represent the process of cortical segmentation and parcellation via the FreeSurfer software suite[21]. The 'FreeSurfer analysis' term has logical axioms describing its required input of a high resolution T1w image and its outputs of (sub-)cortical segmentation maps and (sub-)cortical volume measurement data. The outputs both have logical axioms linking them to all 86 regions of the FreeSurfer atlas that they are about via URIs from Uberon[22].

Figure 2 provides an overview of the 'image data set analysis' process.

3. Use Cases

In combination with the SPARQL Protocol and RDF Query Language (SPARQL) [[23]] and tools that tie into the DICOM framework (i.e. pydicom), we have leveraged the 'image data set' class and MRIO to work with neuroimaging data is several use cases. These use cases come from an automated neuroinformatics platform that works on large public datasets and real-world imaging data.

3.1. Automated MRI Scan Classification

The first use case to arise from MRIO was the automation of scan classification – the process of identifying the acquisition type of an MRI from its DICOM header. This is an important aspect of working with neuroimaging data because some analyses can only be performed on specific kinds of MR images (e.g. segmentation of brain structures using anatomical images vs. mapping the tracts that connection brain regions using diffusion weighted imaging). Information pertaining to the type of MR image may be inferred from acquisition parameters present in the DICOM header. However, this has historically been a difficult problem to solve, as scan type is not directly encoded in DICOM header tags, and different scanner manufacturers use slightly different acquisition parameters for the same scan type.

As previously presented at ICBO 2019, MRIO uses a range of values for common acquisition parameters to address this, using the HermiT reasoner to automatically infer scan type. We have expanded this functionality by using pydicom to automatically parse useful acquisition parameters from DICOM headers and then using SPARQL/Update queries to add new scans as individuals, followed by calling ROBOT [[24]] with the HermiT reasoner to infer the type of the new 'image data set' individuals. The end result is a fully automated method for classifying MRI data without much input or experience from the user. However, as was pointed out in the initial ICBO 2019 presentation of MRIO, inference with HermiT is slow on consumer hardware.

A much higher level of performance may be obtained using machine learning algorithms, such as XGBoost [[25]], trained on the same acquisition parameters used by the 'MR image data set' subclasses from MRIO. The current iteration is a model that predicts scan types using pydicom similar to the previous approach, and outputs the predicted URI of the 'MR image data set' class. The model classifies MRI acquisition types based on features derived from MRIO with 99.953% accuracy, and a macro-averages F1=0.8743. This output can then be chained to other SPARQL queries for further automation of the neuroimaging research process.

3.2. Automated Assignment of Analyses to MRI data

In order to extract useful information from MRI data, the next step to automate was the determination of applicable analysis types to the now classified 'MR image data sets.' This was done using the RDFLib Python library [[26]] in conjunction with a function that processes strings from user input into a SPARQL query to search MRIO for all 'image data set analyses' that accept the specified 'MR image data set' types as input. These outputs are then used to coordinate the pipeline engine of the neuroinformatics platform and direct storage of the data derived from analyses.

Using a Python function to generate the SPARQL queries from user input provides a great deal of modularity and extends this feature's use. The end result is both a script that may be called by a user who simply wants to know what programs may be used to analyze their imaging data and a tool that powers a sophisticated platform that carries out the analyses assigned by the SPARQL query. Figure 3 contains an example of the script available for end users.

3.3. Automated Transformation of Data into Standard Formats

Because MRIO focuses on neuroimaging with MRI, there is potential to integrate with the BIDS specification. BIDS is a standard format for organizing neuroimaging datasets and the data that may be derived from them. This organization is typically hierarchical and follows prescribed naming conventions to keep annotated MR images consistent across datasets. Although developed to be practical rather than ontological, the naming conventions used in BIDS mostly align with the 'MR image data set' subclasses (e.g. 'T2 FLAIR image data set' -> 'sub-id_FLAIR.nii.gz/json'). An additional benefit of BIDS is the ability to use "BIDS-apps," containerized neuroimaging programs that expect only a valid BIDS dataset as input. These BIDS-apps typically come preconfigured with sane defaults for a one-size-fits-all approach for working with neuroimaging data, greatly alleviating the burden of writing up processing pipeline scripts, cleaning data, and organizing results for researchers. Because BIDS-apps are typically containerized, methods and data can be kept consistent from one dataset or study to the next. These factors combine to make BIDS a powerful tool for harmonizing large imaging datasets.

One of the largest barriers to the widespread adoption of BIDS is the investment required to transform neuroimaging data into valid BIDS datasets, in terms of both time and effort. It is necessary to know the scan types of the MRI data at the outset, and the process of generating a BIDS dataset typically requires scripting and manual annotation of data. BIDS expects images to be in the Neuroimaging Informatics Technology Initiative (NIfTI) [27] file format commonly used in research settings with header data moved to accompanying JSON files, rather than DICOM. While DICOM provides standardized reporting for the acquisition of individual scans, a typical scanning session will be dumped into directories according to machine-readable IDs that are unintelligible to the average researcher. Therefore, creating a valid BIDS dataset from DICOM requires specialized software and domain expertise. Fortunately, a software tool called dcm2niix [28] is capable to converting DICOMs into the NIfTI format required by BIDS, while also automatically parsing data from DICOM headers and generating the JSON files for BIDS. Using the dcm2niix tool in conjunction with the automated 'MR image data set' type classification from MRIO, it is possible to automatically generate valid BIDS datasets from an unsorted DICOM directory (Figure 4), greatly reducing the time and effort and allowing researchers to easily process their data using BIDS-apps.

4. Discussion

Our work using the 'image data set' and 'image data set analysis' classes helps harmonize large datasets in biomedical imaging in several ways. It is possible to automate MRI scan classification using either the HermiT reasoner or a machine learning model to predict the types of 'MR image data set' present in datasets of unsorted DICOMS. It is then possible to automatically assign programs to analyze the newly annotated 'MR image data sets' according to the relevant 'image data set analysis' classes. Furthermore, these 'MR image data set' classes can help transform unsorted DICOM directories into standard formats such as the BIDS specification.

In addition, our inclusion of 'brain region atlas image data sets' provide a template for canonical images of the human body and its anatomy, allowing for annotation of biomedical imaging data according to the body part scanned. We have used the Uberon [22] anatomy ontology in our own work to annotate the results of an 'image data set analysis' according to the URI of the anatomical brain region they pertain to.

We have also developed an 'image segmentation map' subclass that can be used for annotating datasets used for machine learning and computer vision. This utility extends biomedical imaging and reflects the term's general usability.

4.1. Limitations

The 'image data set' class, its children, and the 'image data set analysis' class all originate from MRIO, which is an application ontology focused primarily on neuroimaging in MRI. As such, future work will need to be done to include additional biomedical imaging modalities such as CT and PET. However, the high level 'image data set' class is broad enough that these can easily be added.

Additionally, there is an editor's note to 'data set,' the parent class to 'image data set' that states:

2009/10/23 Alan Ruttenberg. The intention is *that this term represent collections of like data. So this isn't for, e.g. the whole contents of a cel file, which includes parameters, metadata etc. This is more like java arrays of a certain rather specific type*

However, it can be argued that the data captured in a DICOM is a collection of like data all describing the same process of acquiring a biomedical image. Or it may simply be that 'image data set' belongs under 'data item' (IAO:0000027) rather than 'data set.' This is an issue that demands further discussion with the OBI Consortium to resolve.

5. Conclusion

Here we presented the 'image data set' class, along with the 'image data set analysis' class, which we have developed to capture the processes of acquisition, annotation, and analysis of biomedical imaging data in an effort to better harness the vast amount of untapped imaging datasets. We also demonstrated several ways we have been using the 'image data set' class with MRIO to facilitate our work with large public data sets and real world imaging data.

The ontology and related scripts are publicly available with CC-BY 4.0 licensing at https://github.com/Buffalo-Ontology-Group/MRI_Ontology.

6. Acknowledgements

AD (A. 6) was supported by 5UL1TR001412 (NCATS).

MD (A.5) has received consultant fees from Claret Medical and EMD Serono, and research grant support from Novartis and Celgene.

This manuscript has been improved by the insightful comments of the anonymous reviewers.

7. Figures and Tables

Table 1
Higher-level terms

Label	Definition
'image data set'	A data set that is comprised of structured measurements about some entity and its associated metadata using pixels (2D), voxels (3D), or an arbitrary number of dimensions. An image data set can be the source from which an image is produced.
'raw image data set'	An image data set that encodes measurement values produced by some instrument before undergoing a data transformation.
'computed image data set'	An image data set that is the output of an image data set analysis.
'brain region atlas image data set'	An image data set consisting of values computed from multiple image data sets encoded to represent the spatial location of individual functional or structural regions of a canonical brain.
'image segmentation map'	An image data set of integer values in which each value corresponds to some shared characteristic or computed property. The values often belong to a group of pixels or voxels that share the same characteristic, such as a tissue type or anatomical region.
'image data set analysis'	The process of deriving a data item from an image data set using computer algorithms. The produced data item can be an image data set, data measurement, or any other data item.

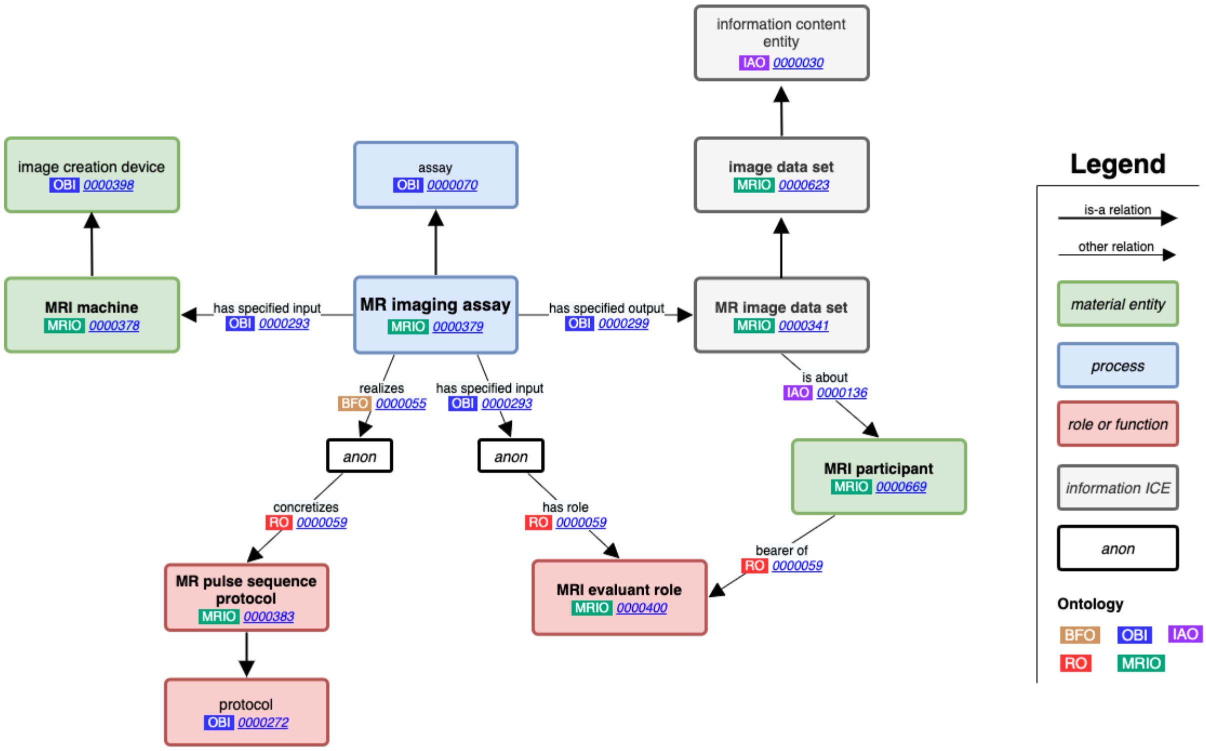

Figure 1: An overview of the acquisition of a 'magnetic resonance image data set.'

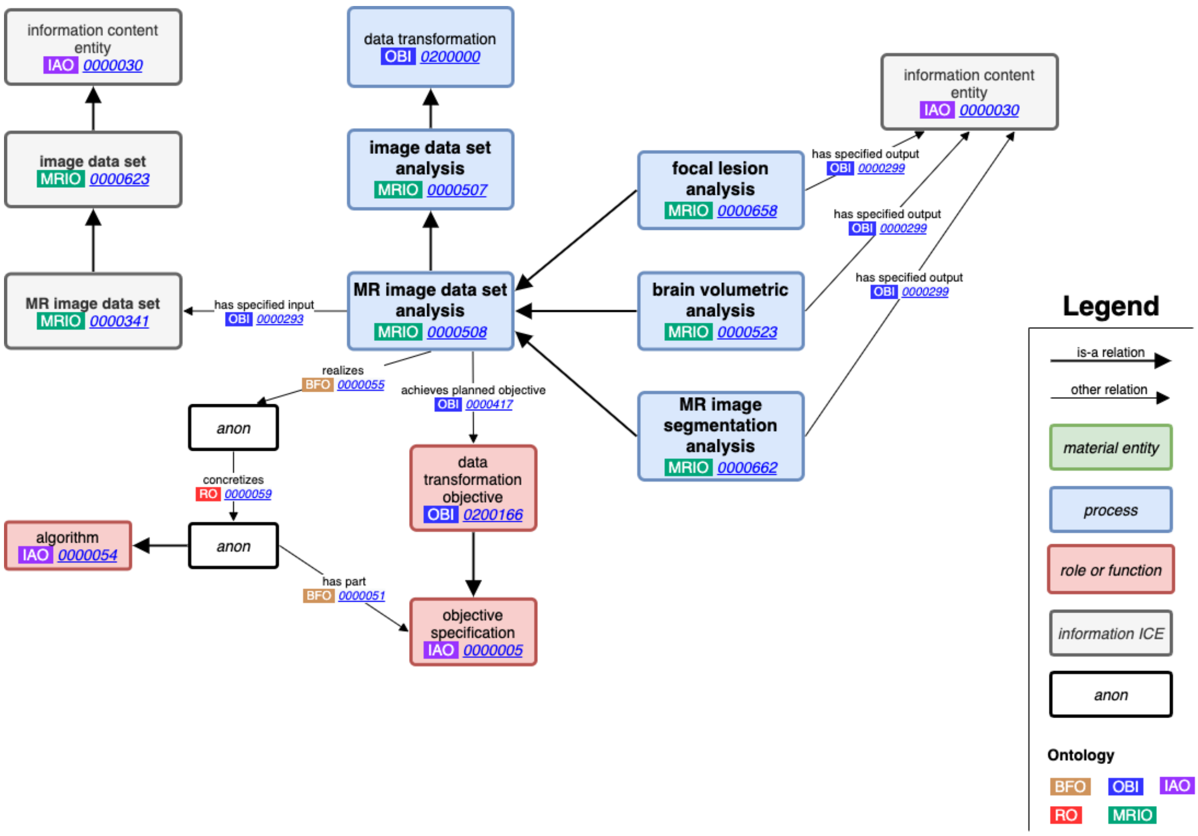

Figure 2: An overview of the 'image data set analysis' process.

```
63 find_analyses(['T1 weighted', 'FLAIR'])

{'analyses': ['http://purl.obolibrary.org/obo/MRIO_0000525',
  'http://purl.obolibrary.org/obo/MRIO_0000515',
  'http://purl.obolibrary.org/obo/MRIO_0000676',
  'http://purl.obolibrary.org/obo/MRIO_0000514',
  'http://purl.obolibrary.org/obo/MRIO_0000509',
  'http://purl.obolibrary.org/obo/MRIO_0000522',
  'http://purl.obolibrary.org/obo/MRIO_0000517',
  'http://purl.obolibrary.org/obo/MRIO_0000656',
  'http://purl.obolibrary.org/obo/MRIO_0000525'],
 'labels': ['SIENAX analysis',
  'FreeSurfer Analysis',
  'FAST Analysis',
  'FIRST analysis',
  'DeepGRAI analysis',
  'LPA analysis',
  'NeuroSTREAM analysis',
  'Deep roboNAC analysis',
  'SIENAX analysis']}
```

Figure 3: An example of the automate analysis assignment script

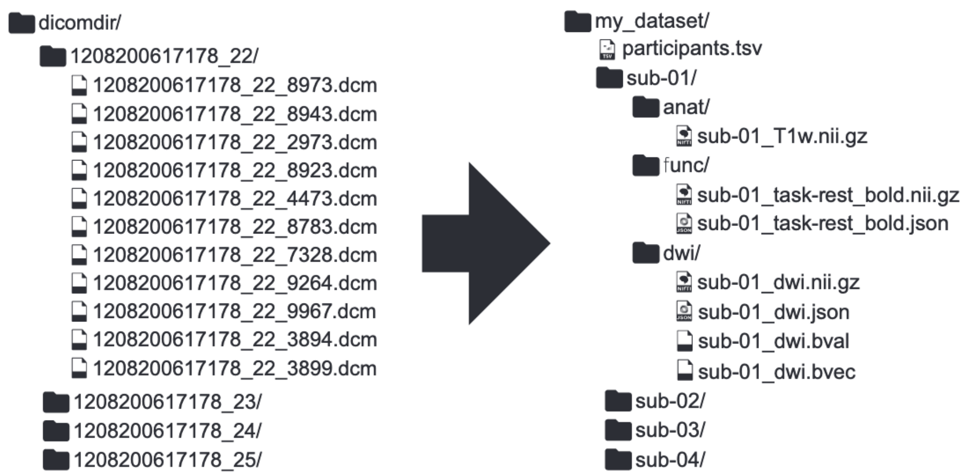

Figure 4. An example transformation of an unsorted DICOM directory into a BIDS dataset

8. References

1. Robb RA. *Three-Dimensional Biomedical Imaging*. Vol 2. CRC Press Boca Raton, FL; 1985.
2. Smith-Bindman R, Kwan ML, Marlow EC, et al. Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016. *JAMA - J Am Med Assoc*. 2019;322(9):843-856. doi:10.1001/jama.2019.11456
3. Sistrom CL, McKay NL. Costs, charges, and revenues for hospital diagnostic imaging procedures: Differences by modality and hospital characteristics. *J Am Coll Radiol*. 2005;2(6):511-519. doi:10.1016/j.jacr.2004.09.013
4. Dwyer M, Carolus K, Bergsland N, et al. Deep learning enables thalamic atrophy measurement on clinical quality T2 FLAIR images (P5.2-057). *Neurology*. 2019;92(15 Supplement):P5.2-057.
5. Fuchs TA, Dwyer MG, Jakimovski D, et al. Quantifying disease pathology and predicting disease progression in multiple sclerosis with only clinical routine T2-FLAIR MRI. *NeuroImage Clin*. 2021;31:102705. doi:10.1016/j.nicl.2021.102705
6. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Jpn J Radiol*. 2019;37(1):73-80. doi:10.1007/s11604-018-0796-2
7. Adduru VR. Automated brain segmentation methods for clinical quality MRI and CT images. *Diss Abstr Int Sect B Sci Eng*. 2020;81(12-B). http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2020-51430-224&site=ehost-live
8. Levine AB, Peng J, Farnell D, et al. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. *J Pathol*. 2020;252(2):178-188. doi:https://doi.org/10.1002/path.5509

9. Bratke G, Rau R, Kabbasch C, Maintz D, Haneder S, Weiss K. Accelerated Imaging of the Lumbar Spine: Compressed Sensing for the Clinical Routine. *Semin Musculoskelet Radiol*. 2019;23(S 01):A008.
10. Mildenberger P, Eichelberg M, Martin E. Introduction to the DICOM standard. *Eur Radiol*. 2002;12(4):920-927. doi:10.1007/s003300101100
11. Mason D, scaramallion, mrbean-bremen, et al. pydicom/pydicom: pydicom 2.3.0. Published online March 29, 2022. doi:10.5281/ZENODO.6394735
12. Trispark Java DICOM Toolkit. Published 2020. https://www.javadicomtoolkit.com/
13. McCormick M, Liu X, Ibanez L, Jomier J, Marion C. ITK: enabling reproducible research and open science . *Front Neuroinformatics* . 2014;8. https://www.frontiersin.org/article/10.3389/fninf.2014.00013
14. Gorgolewski KJ, Auer T, Calhoun VD, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*. 2016;3:160044. doi:10.1038/sdata.2016.44
15. Smith B, Ceusters W. Aboutness: Towards foundations for the information artifact ontology. *CEUR Workshop Proc*. 2015;1515:1-5.
16. Kumar A, Smith B. Oncology Ontology in the NCI Thesaurus BT - Artificial Intelligence in Medicine. In: Miksch S, Hunter J, Keravnou ET, eds. Springer Berlin Heidelberg; 2005:213-220.
17. Serra L. *Early Steps of an Ontology for Magnetic Resonance Imaging: MRIO*.; 2019.
18. Bandrowski A, Brinkman R, Brochhausen M, et al. The Ontology for Biomedical Investigations. *PLoS One*. 2016;11(4):e0154556. doi:10.1371/journal.pone.0154556
19. Noy NF, Crub M, Fergerson RW, et al. Protege-2000 : An Open-Source Ontology-Development and Knowledge-Acquisition Environment. Published online 2003:2003.
20. Glimm B, Horrocks I, Motik B, Stoilos G, Wang Z. HermiT: an OWL 2 reasoner. *J Autom Reason*. 2014;53(3):245-269.
21. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage*. 1999;9(2):195-207.
22. Mungall CJ, Torniai C, Gkoutos G V, Lewis SE, Haendel MA. Uberon, an integrative multi-species anatomy ontology. *Genome Biol*. 2012;13(1):R5. doi:10.1186/gb-2012-13-1-r5
23. Seaborne A, Prud'hommeaux E. SPARQL query language for RDF. *W3C Recomm (January 2008)*. Published online 2008.
24. Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA. ROBOT: a tool for automating ontology workflows. *BMC Bioinformatics*. 2019;20(1):1-10.
25. Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting. *R Packag version 04-2*. 2015;1(4):1-4.
26. RDFLib. Published 2022. https://rdflib.dev/
27. Neuroimaging Informatics Technology Initiative. Published 2013. https://nifti.nimh.nih.gov/
28. Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *J Neurosci Methods*. 2016;264:47-56. doi:10.1016/j.jneumeth.2016.03.001