DEVELOPMENT OF AN IMAGE DATA SET CLASS

Its Role in Biomedical Imaging and Neuroimaging Research

Jacobs School of Medicine and Biomedical Sciences
University at Buffalo

BNAC
Buffalo Neuroimaging Analysis Center

https://github.com/Buffalo-Ontology-Group/MRI_Ontology

Outline

- Introduction
  - Background on biomedical imaging
  - The DICOM standard
  - Ontological representation of imaging data
- Development
  - image data set
  - image data set analysis
- Use cases
  - MRI scan classification
  - Assignment of analyses to MRI data
  - Querying a relational database
  - Transformation into standardized formats
- Conclusion

INTRODUCTION

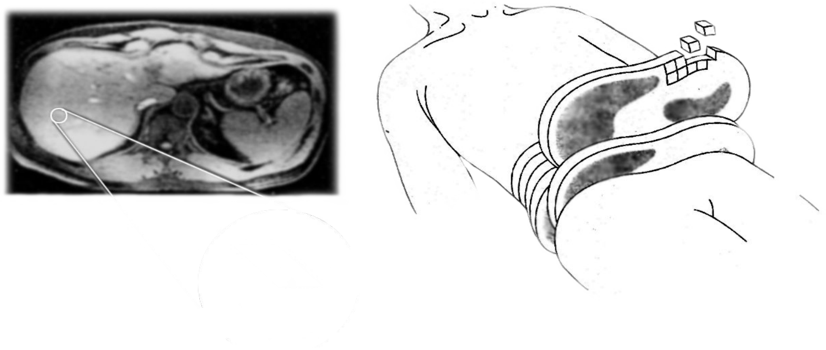
Background on biomedical imaging

The DICOM standard

Ontological representation of imaging data

Biomedical imaging

- Method of visualizing a subject's internal structural and functional anatomy
- Has been a mainstay of modern medicine for decades (Robb RA 1985)
- Includes CT, PET, **MRI**
- The number of biomedical imaging scans performed clinically increases year-over-year (Smith-Bindman R, et al 2016)



Clinical vs Research Imaging

Clinical routine

- Typically lower quality
- Used to guide treatment
- Potentially huge amounts of data generated, but left in silos

Research

- Typically higher quality
- Used to ascertain novel truths about the body
- Lower potential for large datasets

Use of Clinical Images

- Imaging is very expensive (Sistrom CL, et al 2005)
- Push in recent years to mobilize the large quantity of images acquired via clinical routine for use in research, but clinical imaging data is often left unorganized and locked down (Dwyer MG, et al 2019; Fuchs TA, et al 2021)
- There exists then a need for a standardized method of automatically sorting and annotating these large amounts of imaging data
- Would allow researchers to more easily analyze and share data.

The DICOM standard

"DICOM® is the international standard to transmit, store, retrieve, print, process, and display medical imaging information"

- From https://www.dicomstandard.org/, (Mildenberger et al, 2002)

- ISO recognized digital format for biomedical imaging
- Allows for interoperability across modalities, scanner manufacturers, and healthcare systems
- Encodes data pertaining to image acquisition (e.g. date, time, patient age, acquisition parameters, modality, etc.) in the file headers
- Large ecosystem of software and libraries for programmatically working with DICOMs

BIDS

- The field of neuroimaging in particular has recently been moving towards the adoption of the Brain image data set (BIDS) specification (Gorgolewski, KJ, et al 2016), which is a prescribed format for naming and laying out directories of neuroimaging data
- 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

Ontological Representation of Imaging Data

- Much of the data encoded in DICOM headers are semantic strings that often differ between institutions and manufacturers
- Important information for sorting and analyzing data are often disparate from one data set to the next, and different methods must be used to work with different data sets
- High barrier of entry for working with imaging data, requiring experience in fields like computer science and informatics that clinical researchers might not have
- Biomedical imaging data may be represented using high-level classes that integrate with DICOM and BIDS to help harmonize large amounts of existing data

DEVELOPMENT

image data set

image data set analysis

MRIO at ICBO 2019

- Development on MRIO began by Dr. Serra and Dr. Diehl in collaboration with Dr. Dwyer, presented at ICBO 2019
- Made decision to construct top-down and bottom-up using OBO Foundry principles
- Added ~70 new terms
- Uses BFO as base, extends OBI and IAO
- Intended to provide formal representation of MRI acquisition, a means of querying MRI data, and automated scan type classification using data from DICOM headers

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.

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

Serra et al, 2019

image data set

Definition - 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.

Contributing to OBI (OBI:0003327)

image data set analysis

Definition - 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.

Contributing to OBI (OBI:0003355)

‘image data set’ vs image

Why not image (IAO_0000101)?

- image is specifically defined as two dimensional, but MRI and many biomedical imaging techniques typically encode data pertaining to three or four dimensions
- image is a subclass of figure, which emphasizes the importance of some two dimensional *arrangement* of ICE’s relating information to an entity

Thus, we have a need for a more general term for an information content entity that more fully represents the different kinds of information found in an MRI scan.

The definition for data set states that it is a collection of data items (e.g. scan, scanner settings, subject data, etc.) that have something in common (i.e. the participant). We felt this sufficient for the parent class of image data set.

Use of `image data set`

We believe that a general `image data set` class has utility beyond the domain of MRI/neuroimaging

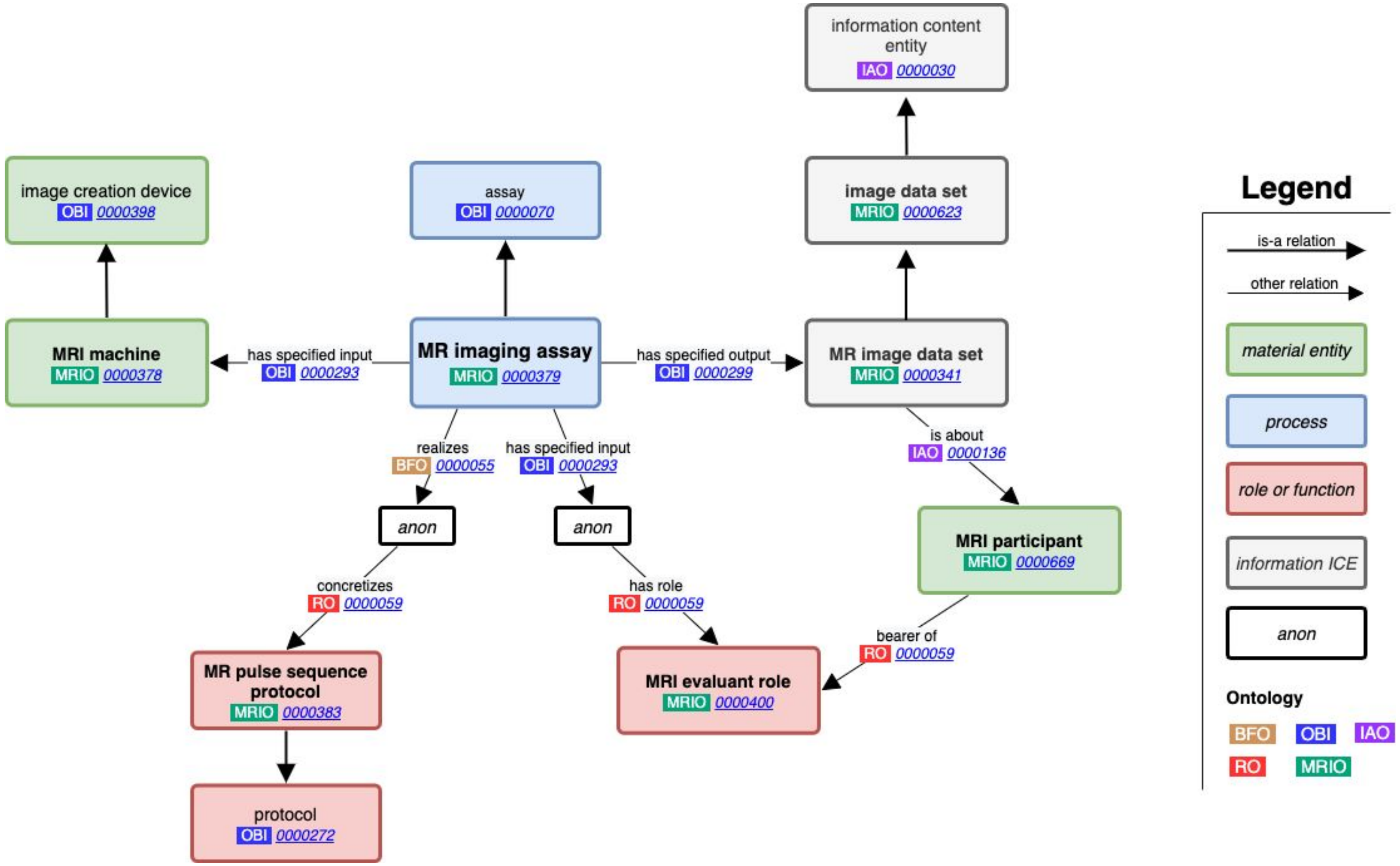
- Applies to any biomedical images in DICOM format (CT, PET, etc.)
- Flexibility in n-dimensions allows for potential use beyond static, two dimensional images
  - Animations, time series, etc.
- Many digital image file formats contain headers bearing additional information (e.g. date, location, color space, etc.) that aren't adequately captured in the `image` class

Use of `image data set`

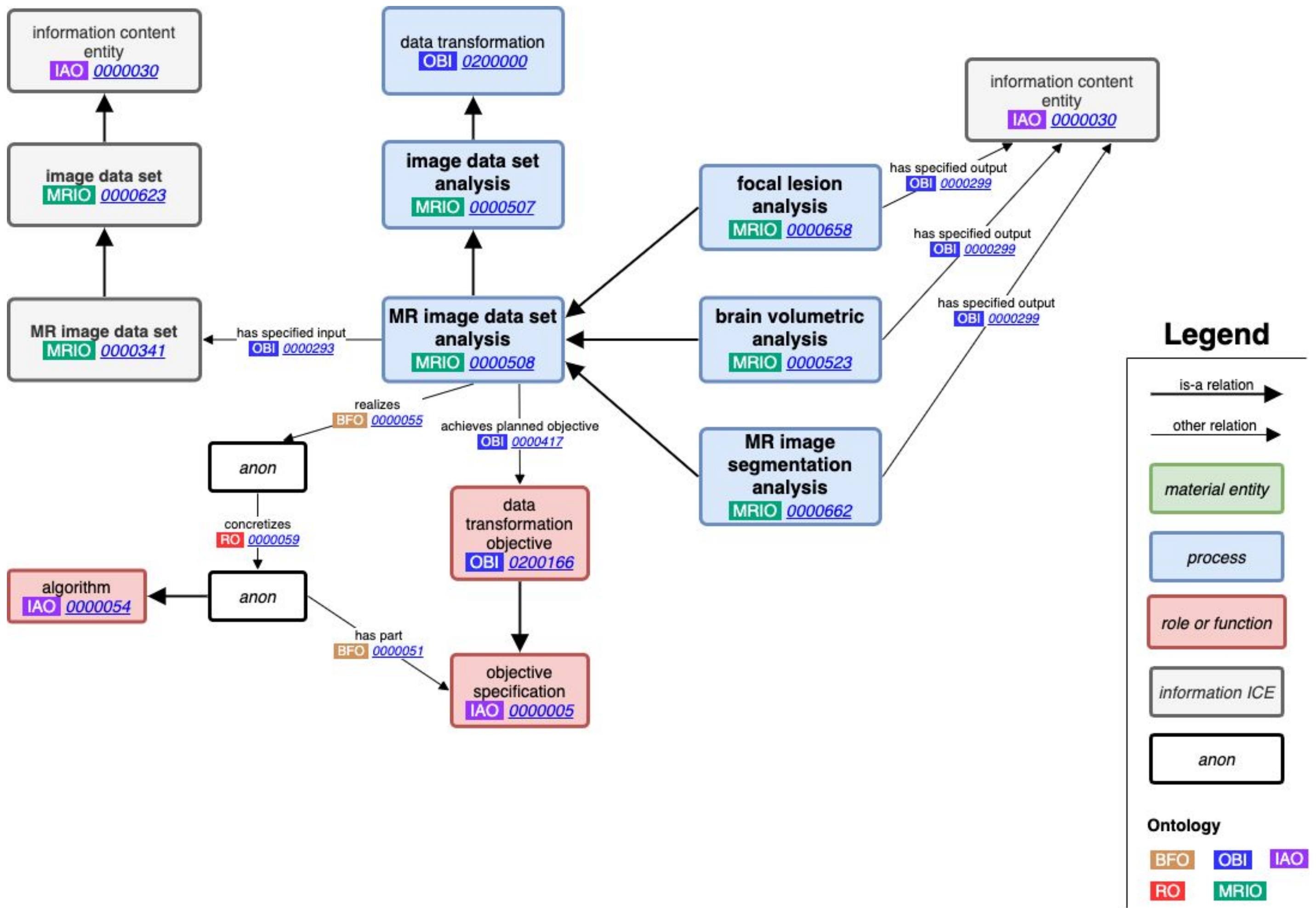
- planned process
  - assay
    - **MR imaging assay**
      - **diffusion weighted MRI acquisition sequence**
      - **functional MRI acquisition sequence**
      - **Proton density sequence**
      - **T1 weighted MRI acquisition sequence**
        - **T1 GRE sequence**
          - **2D T1 GRE sequence**
          - **3D T1 GRE sequence**
        - **T1 SE sequence**
      - **T2 weighted MRI acquisition sequence**
        - **T2 FLAIR sequence**
        - **T2 FSE sequence**
        - **T2 STIR sequence**
        - **T2* GRE sequence**
    - volume measurement assay

- information content entity
  - data item
    - data about an ontology part
    - data set
    - measurement datum
    - **network graph quality datum**
    - setting datum
      - **MRI machine setting datum**
        - **MRI machine parameter**
          - **acquisition matrix**
          - **diffusion b value**
          - **diffusion gradient**
          - **echo time**
          - **echo train length**
          - **field of view**
          - **flip angle**
          - **inversion time**
          - **matrix size**
          - **number of temporal positions**
          - **phase encoding**
          - **phase FOV**
          - **plane**
            - **coronal**
            - **sagittal**
            - **transverse**
          - **repitition time**
          - **slice gap**
          - **slice thickness**

Magnetic Resonance Imaging Assay



Magnetic Resonance image data set Analysis



USE CASES

MRI scan classification

Automated assignment of analyses for all scans in a session

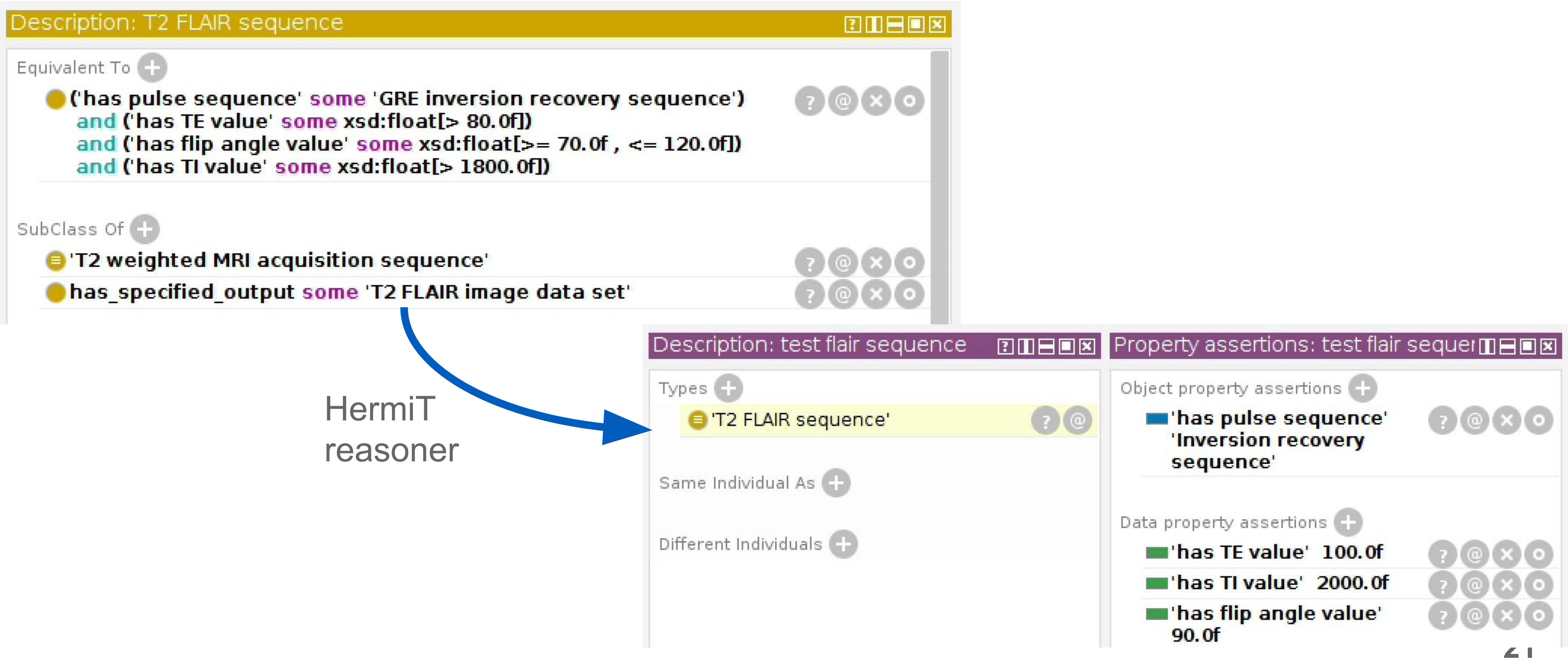
Querying a relational database

Integration with BIDS spec

Use in Neuroinformatics Platform

Unsorted DICOMs	
Automated classification	(XGBoost model trained using parameters chosen form MRIO)
Imaging analysis	
Automated assignment of analyses	(MRIO and SPARQL queries)
Provide informatics to clinical researchers	
Intelligent queries on annotated results of analyses	(Ontop)

Automated MRI scan classification



Use of MRIO/SPARQL to assign analyses

Scenario:

1. A researcher collects MRI scans as part of a larger study (T1 and T2 FLAIR for each subject)
2. They want to know what kinds of analyses they can do with these scans and what the results mean
3. The researcher can query MRIO and find all the pertinent analyses for T1 weighted and T2 FLAIR scans, as well as further information on the types of analyses and results

```
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']}
```

Credit: Mackenzie Smith for developing the SPARQL query and function

Current state hashes class labels to MRIO IRIs for each scan type

Potential for robust querying system using NLP to map user inputs

Ontop* and XNAT** Database Integration

```
PREFIX obo: <http://purl.obolibrary.org/obo/>

SELECT ?scan_id ?measurement_value
WHERE {
  ?scan_id obo:IAO_0000136 obo:UBERON_0010225 ;
           obo:IAO_0000004 ?measurement_value .
}
```

Simple English:

- "Find anything that is about some thalamus, and has some measurement value"

Because '*thalamic volume measurement datum*' 'is about' some *thalamus*, results of thalamic volumetry analyses are returned for this query

*Calvanese et al, 2016

**Marcus et al, 2007

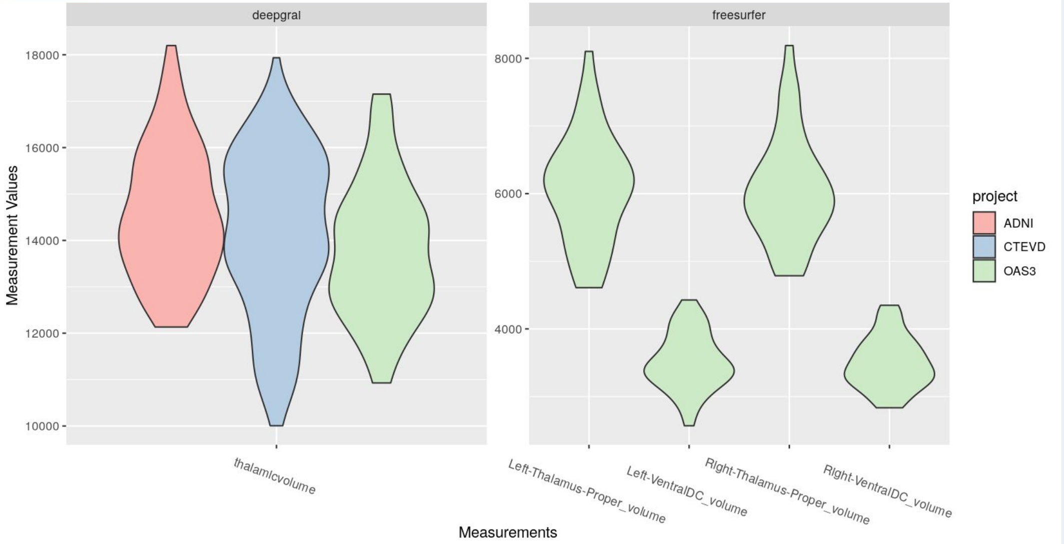
Jacobs School of Medicine and Biomedical Sciences
University at Buffalo

Use in Neuroinformatics Platform

ROI

Thalamus

Submit



CSV Excel

Search:

	ID	label	age	gender	race	project	deepgrai_thalamicvolume	freesurfer_Left-VentralDC_volume	freesurfer_Right-Thalamus-Proper_volume	freesurfer_Left-Thalamus-Proper_volume	freesurfer VentralDC_
195	CBI_XNAT13_E00001	test_MR_1				CBI	16646.9609375				
194	CBI_XNAT24_E02160	OAS30491_MR_d0074	78	female		OAS3	13084.296875				
193	CBI_XNAT19_E00025	OAS30482_MR_d1408	54	female		OAS3	14367.4375	3612.9	6529.5	6838.8	
192	CBI_XNAT24_E02495	OAS30479_MR_d2421	86	male		OAS3	13351.6416015625				
191	CBI_XNAT24_E04082	OAS30475_MR_d0062	72	female		OAS3	12573.1669921875				
190	CBI_XNAT19_E00024	OAS30474_MR_d0069	78	female		OAS3		3443.7	6184.6	6157.4	
189	CBI_XNAT24_E02210	OAS30464_MR_d2848	63	female		OAS3	15233.0859375				
188	CBI_XNAT24_E03986	OAS30445_MR_d0133	65	female		OAS3	14162.212890625				
187	CBI_XNAT24_E02157	OAS30429_MR_d0055	48	female		OAS3	14560.3896484375				
186	CBI_XNAT21_E00004	OAS30422_MR_d0104	76	male		OAS3	13019.6796875				

Showing 1 to 10 of 377 entries

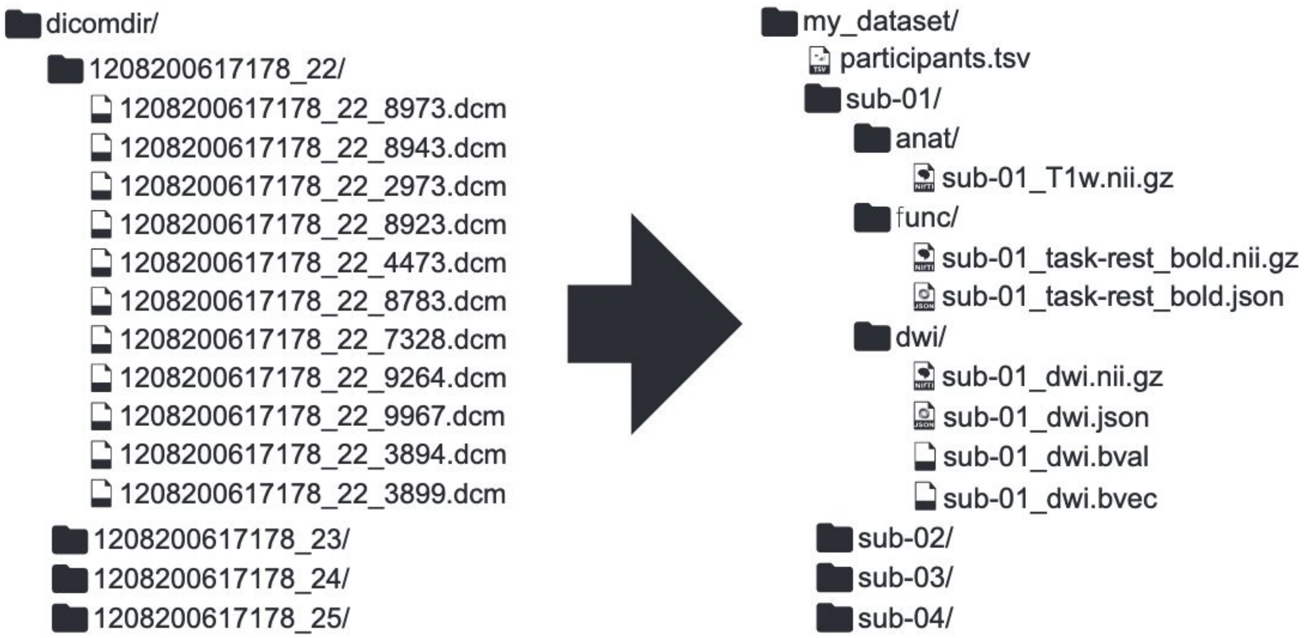
Previous 1 2 3 4 5 … 38 Next

Brain Imaging Data Structure, or BIDS

- Standard specification for storing and organizing neuroimaging data
  - Mostly MRI, but also supports PET, MR spec, EEG
- Sort data according to scan type (similar to MRIO)
- One goal is to use MRIO and SPARQL queries to generate a BIDS dataset from an unsorted DICOM directory
  1. Classify scans
  2. Map MRIO labels to BIDS folder names
  3. String processing to rename and move files

```
sub-control01/
    anat/
        sub-control01_T1w.nii.gz
        sub-control01_T1w.json
        sub-control01_T2w.nii.gz
        sub-control01_T2w.json
    func/
        sub-control01_task-nback_bold.nii.gz
        sub-control01_task-nback_bold.json
        sub-control01_task-nback_events.tsv
        sub-control01_task-nback_physio.tsv.gz
        sub-control01_task-nback_physio.json
        sub-control01_task-nback_sbref.nii.gz
    dwi/
        sub-control01_dwi.nii.gz
        sub-control01_dwi.bval
        sub-control01_dwi.bvec
    fmap/
        sub-control01_phasediff.nii.gz
        sub-control01_phasediff.json
        sub-control01_magnitude1.nii.gz
        sub-control01_scans.tsv
code/
    deface.py
derivatives/
README
participants.tsv
dataset_description.json
CHANGES
```

Use of MRIO/SPARQL to transform DICOM -> BIDS



Conclusion

Our work using the ‘image data set’ and ‘image data set analysis’ classes helps harmonize large datasets in biomedical imaging

- Automated scan classification and assignment of analyses makes it easier to work with large, unsorted MRI datasets
- Integration with DICOM and BIDS helps clinical and translational researchers leverage neuroimaging data

Most of the development has been focused on MRI, but ‘image data set’ is a high-level term that may also be used with other imaging modalities (e.g. PET, CT, etc.)

Note: Currently working on integrating with OBI, suggestion to change to ‘image data set’

Acknowledgements

Lucas Serra

Mackenzie Smith

William Duncan

Alan Ruttenberg

Lauren Wishnie

Michael Dwyer

Alexander Diehl

Work cited

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. Mildenberger, P., Eichelberg, M., & Martin, E. (2002). Introduction to the DICOM standard. *European Radiology*, *12*(4), 920–927. https://doi.org/10.1007/s003300101100
7. Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44
8. Serra, L. (2019). *Early steps of an Ontology for Magnetic Resonance Imaging: MRIO*.
9. Calvanese, D., Cogrel, B., Komla-Ebri, S., Kontchakov, R., Lanti, D., Rezk, M., Rodriguez-Muro, M., Xiao, G., & Corcho, Ó. (2016). Ontop: Answering SPARQL Queries over Relational Databases. *Semantic Web*, *8*(3), 471–487.
10. Marcus, D. S., Olsen, T. R., Ramaratnam, M., & Buckner, R. L. (2007). The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, *5*(1), 11–34. https://doi.org/10.1385/ni:5:1:11