



ReproNim: A Center for Reproducible Neuroimaging Computation

<http://www.repronim.org>

Course Material

Screenshot of a GitHub repository page for "Repronim / sfn2018-training".

The repository details:

- Owner: ReproNim
- Name: sfn2018-training
- Unwatched (8 watchers)
- Starred (2 stars)
- Forked (3 forks)

Repository statistics:

- 103 commits
- 3 branches
- 1 release
- 1 environment
- 10 contributors

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

Commit history (partial list):

Author	Message	Time
jgrethe	Added FAIR data folder	Latest commit 07b71aa 28 seconds ago
.datalad	[DATALAD] new dataset	11 days ago
FAIR	Added FAIR data folder	28 seconds ago
Introduction	added pdf	4 months ago
data	Augmented file with MYSTERY	5 months ago
section1	Fixed commands for Excercise 3	5 months ago
section23	more github urls tune ups	16 days ago
section4		
vm		
.gitattributes	[DATALAD] new dataset	11 days ago
.gitignore	Dump of dumb analysis script for 1st-level FSL GLM	7 months ago
README.md	Strip away aged/duplicate schedule, point to the website	9 days ago
README.md		

https://github.com/Repronim/sfn2018-training



FAIR Data

We will provide an introduction to the FAIR Data Principles and strategies for making research outputs that follow these principles - that data is Findable, Accessible, Interoperable and Re-Usable. This section will focus on making data FAIR via BIDS (Brain Imaging Data Structure) and NIDM (Neuroimaging Data Model) data structures, providing an overview of metadata, common data elements, terminologies/ontologies and semantic annotation of data. This section will also introduce pyNIDM, a Python library to manipulate the Neuro Imaging Data Model.

Section 1

Introduction to FAIR

What is open data?

“Open data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose.”

*Open definition;
Open Knowledge Foundation*

Push for open science-open data

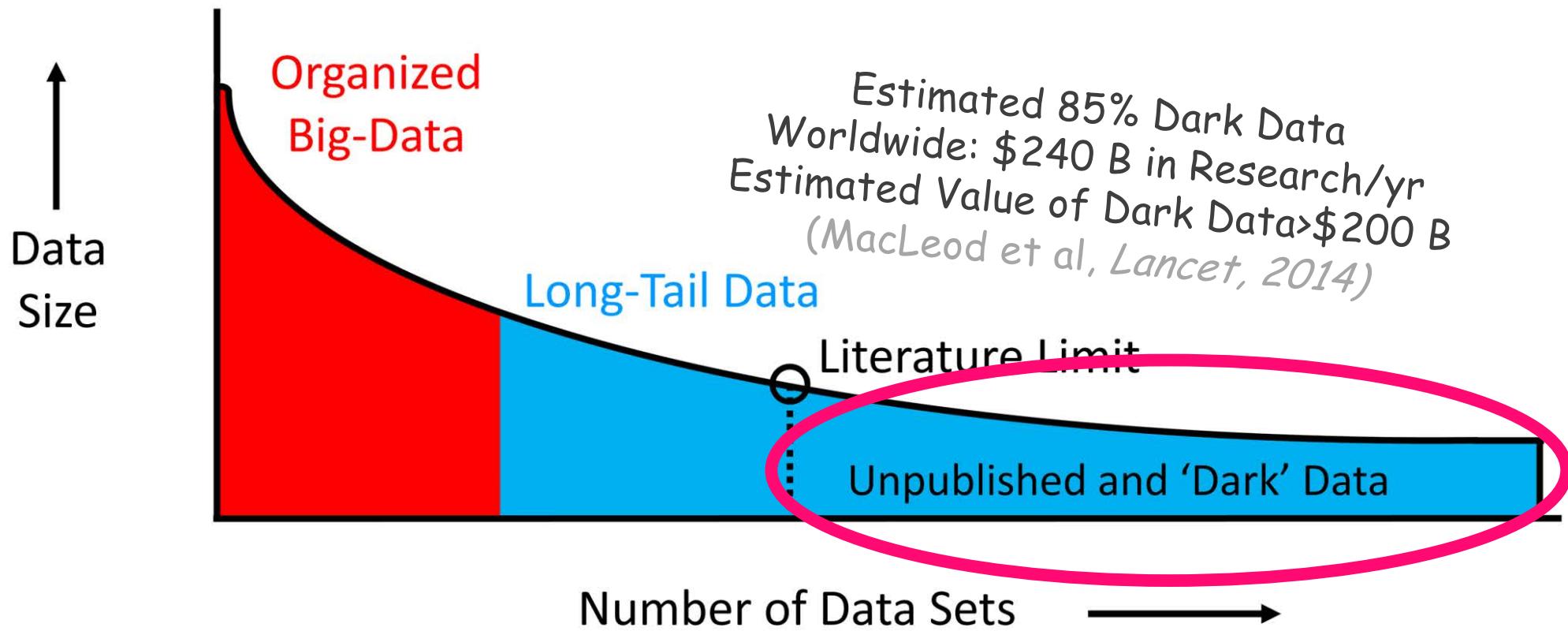
“The Administration is committed to ensuring that, to *the greatest extent* and *with the fewest constraints possible* and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data. “

US Office of Science and Technology Policy
Increasing access to the *results* of federally funded scientific research
Memo, February 22, 2013

Why (publish) data?

- Increase transparency and reproducibility of research
- Allow and encourage re-use of data sets, e.g., by informaticians, researchers, modelers
- Big data from small data: new insights from analysis across data

TOWARD 'DATAFICATION' OF BIOMEDICINE

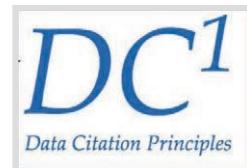


Ferguson et al., 2014, *Nature Neuroscience*

Data as a Research Product

“Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record...”

1. Data should be considered *legitimate, citable products of research*. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.
2. Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.
3. In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.



An Ecosystem for Data Citation

SCIENTIFIC DATA 

Altmetric: 40 [More detail >](#)

Editorial | **OPEN** | Published: 08 May 2018

On the road to robust data citation

Scientific Data 5, Article number: 180095 (2018) | [Download Citation ↓](#)

Scientific Data is changing the way we incorporate links into our data citations. We will now be taking advantage of the resolver services offered by identifiers.org and N2T.net to provide more standardized and predictable links for biomedical datasets that have accession identifiers when they are cited in our publications.

These two parallel services help address an important problem in data citation: many of the most important community-supported biomedical repositories use accession identifier systems that must be resolved according to repository-specific linking rules. Indeed, a single repository may offer multiple link structures that all point to the same dataset. An article published today in the journal describes how identifiers.org and N2T.net, provided by the EMBL-EBI and the California Digital Library respectively, are working together to provide machine-resolvable persistent identifiers for datasets across a wide range of biomedical data repositories¹.

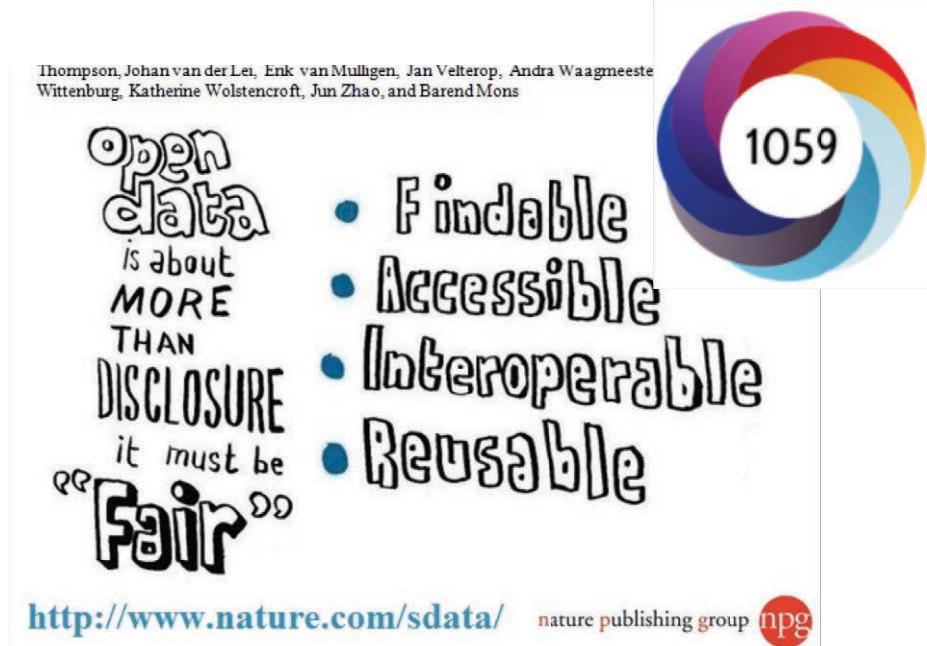


FAIR Principles

SCIENTIFIC DATA

OPEN SUBJECT CATEGORIES » Research data » Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship



- **Findable:** A core aspect of an open data ecosystem - facilitate access to a collection of diverse research resources
- **Accessible:** Assist users by directing them to available resources for that researcher
- **Interoperable:** Provide unified views of certain data and information across resources
- **Reusable:** Supports community standards and develops standard representations for resources
- **Attribution and Recognition:** Initiatives to unambiguously cite resources and data

Assist researchers in ensuring that their resources (e.g. software tools, core facilities,...) and data are FAIR

Making data FAIR

- FAIR data starts with good laboratory management
- Easiest way to make data FAIR is to submit it to a data repository
 - Specialized community repository
 - General repository
 - Institutional repository
- Journals are starting to post lists of repositories that are acceptable (e.g.,
<http://journals.plos.org/plosone/s/data-availability>).
- Data repositories act in the role of publishers for data:
 - links are stable
 - metadata is consistent
 - community standards are adhered to
- Data journals may have more stringent requirements

The cover of the journal **SCIENTIFIC DATA** features a red diagonal stamp with the word "IN PRESS". The title "SCIENTIFIC DATA" is in large blue letters, with "DATA" in a larger font. Below the title is the subtitle "The FAIR Guiding Principles for scientific data management and stewardship". A list of authors follows: Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercé Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hoof, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons.

Open data
is about
MORE
THAN
DISCLOSURE
it must be
“Fair”

- Findable
- Accessible
- Interoperable
- Reusable

<http://www.nature.com/sdata/> nature publishing group npg

ReproNim Training



Data and the FAIR Principles

FAIR is a collection of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable. This module provides a number of lessons to ensure that a researcher's data is properly managed and published to ensure it enables reproducible research.

It is based on the lesson template used in [Neurohackweek](#), [Data Carpentry](#) and [Software Carpentry](#) workshops.

Schedule

09:00	Module Overview: Data and the FAIR Principles	Why is FAIR important? Who is this modulde for? How can I get some help if I get stuck on solving for an exercise or a question? When and where are the future ReproNim training workshops?
09:00	Lesson 1: Introduction to the Web of Data	What is a research object and how do I properly identify it? What is linked open data? What are the FAIR Data principles?
09:02	Lesson 2: Ethics	What ethics policies and issues surround privacy, data sharing, and the use of data?
09:02	Lesson 3: Data Publishing	Am I ready to publish my data? What resources are available for your research data needs?
09:03	Lesson 4: Your Labtory Datastore	What resources are available for me to be a good steward of my laboratory's data
09:03	Lesson 5: Semantic Data Representations	How do I represent my data as linked data?
09:03	Finish	http://www.reproducibleimaging.org/module-FAIR-data/

ReproNim Training

The screenshot shows a web browser window with the address bar containing "reproducibleimaging.org". The main content area displays a lesson titled "Data and the FAIR Principles" under "Lesson 1: Introduction to the Web of Data". The navigation bar at the top includes links for "Home", "Reference", "Episodes", "License", and a search bar. The overall layout is clean and professional.

< Data and the FAIR Principles >

Lesson 1: Introduction to the Web of Data

?

Overview

Teaching: Self Paced min

Exercises: 2 min

Questions

- What is a research object and how do I properly identify it?
- What is linked open data?
- What are the FAIR Data principles?

Objectives

- Understanding identifiers
- History of open and linked data
- Overview of the FAIR principles

Introduction

This lesson provides an overview of strategies for making research outputs available through the web, with an emphasis on data. It introduces concepts such as persistent identifiers, linked data, the semantic web and the FAIR principles. It is designed for those with little to no familiarity with these concepts. More technical discussions can be found in the reference materials.

Lesson Episodes

- Overview of the current ecosystem
- Research objects and identifier systems
- Short history of open & linked data technologies
- Benefits of linked open data with examples
- Towards the FAIR principles

Section 2

Interoperable Data

Motivation for Structured Data

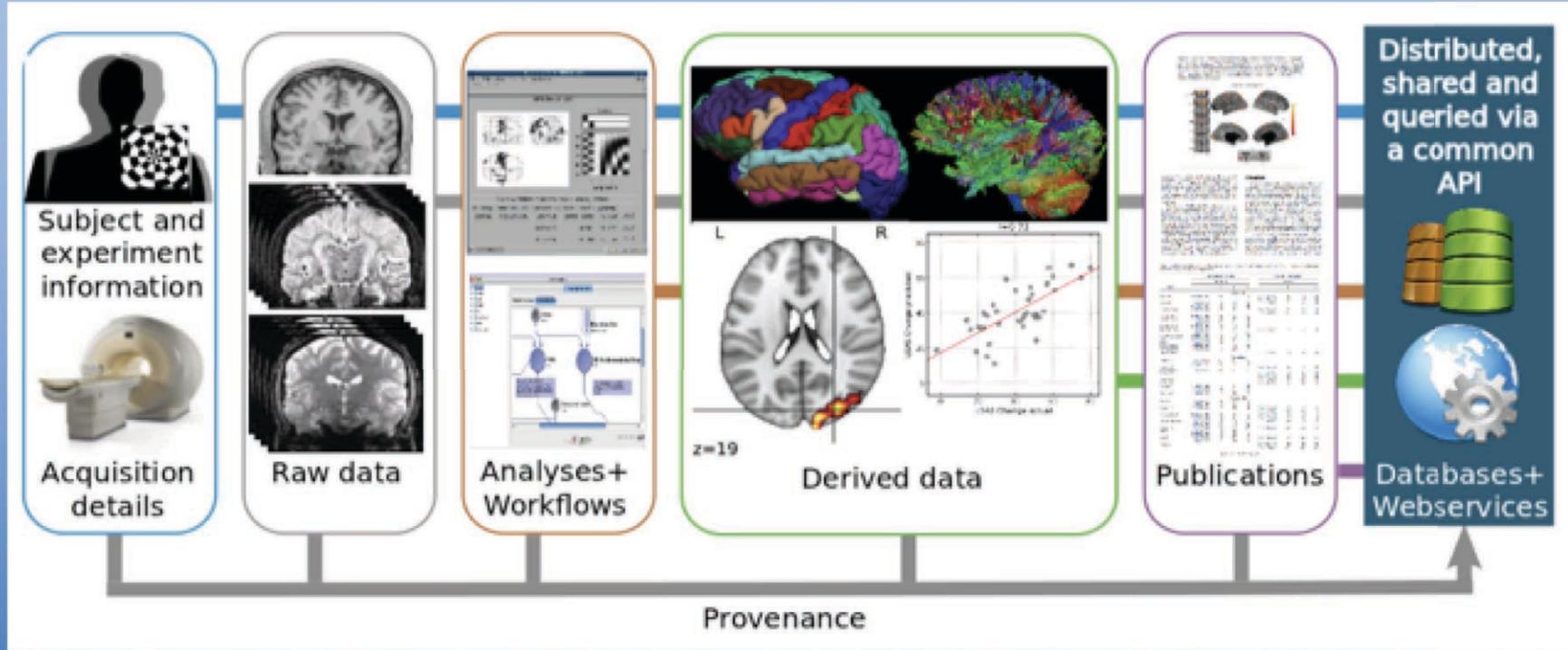
- Imaging experiments are complicated and can be organized in many different ways
- Many files/file types per subject
- Members of the same lab may use different ways to arrange data
- Difficulty sharing data within and across large scale projects



"Think this is bad? You should see the inside of my head."

General Neuroimaging Workflow

Data Flow and Stages of Data Publishing Opportunities



Metadata Descriptions

BIDS Specifications

8.9 Participant file

Template:

(single session case)

`participants.tsv`

`participants.json`

`phenotype/<measurement_tool_name>.tsv`

`phenotype/<measurement_tool_name>.json`

Optional: Yes

The purpose of this file is to describe properties of participants such as age, handedness, sex, etc. In case of single session studies this file has one compulsory column `participant_id` that consists of `sub-<participant_label>`, followed by a list of optional columns describing participants. Each participant needs to be described by one and only one row.

8.9.1 `participants.tsv` example:

<code>participant_id</code>	<code>age</code>	<code>sex</code>	<code>group</code>
<code>sub-control01</code>	34	M	<code>control</code>
<code>sub-control02</code>	12	F	<code>control</code>
<code>sub-patient01</code>	33	F	<code>patient</code>

http://bids.neuroimaging.io/bids_spec.pdf

Participants.tsv

OpenNEURO

PUBLIC DASHBOARD SUPPORT FAQ SIGN IN

Versions < 

Multi-echo fMRI replication sample of autobiographical memory, prospection and theory of mind reasoning tasks -  participants.tsv

PARTICIPANTS.TSV  DOWNLOAD

participant_id	gender	age	physioSampling	restAcquisition
sub-01	M	26	50	after_cuedSGT
sub-02	M	21	50	after_cuedSGT
sub-03	M	22	50	after_cuedSGT
sub-04	M	23	50	after_cuedSGT
sub-05	M	21	50	before_cuedSGT
sub-06	M	19	50	before_cuedSGT
sub-07	F	18	50	before_cuedSGT
sub-08	F	21	50	before_cuedSGT
sub-09	M	20	50	before_cuedSGT
sub-10	F	21	50	before_cuedSGT
sub-11	F	20	50	before_cuedSGT
sub-12	M	21	50	before_cuedSGT
sub-13	F	31	50	before_cuedSGT
sub-14	M	22	50	before_cuedSGT
sub-15	M	22	50	before_cuedSGT
sub-16	M	25	50	before_cuedSGT

Terms versus Common Data Elements

Age 
http://uri.interlex.org/base/ilx_0100400

PreferredId: PATO:0000011 Type: term OWL Equivalent: owl:Class
A time quality inhering in a bearer by virtue of how long it has existed.

Version: 2

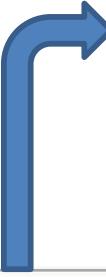
General Children Relationships Annotations Referenced By

Superclasses:  Expand list Age

Synonyms: Synonym

Existing Ids:

Preferred	CURIE	IRI
NLXWIKI:PATO_0000011	http://neurolex.org/wiki/PATO_0000011	
		obo/PATO_0000011
		http://uri.interlex.org/base/ilx_0100400

Log in to suggest term to community 

Back to search results 

Add new term

Export:   

interview_age (Age Differentiation Test) 
http://uri.interlex.org/base/ilx_0345073

PreferredId: ILX:0345073 Type: cde OWL Equivalent: owl:Class
Age in months at the time of the interview/test/sampling/imaging.

Version: 2

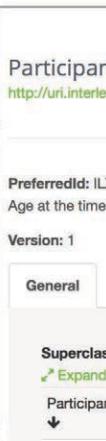
General Children Relationships Annotations Referenced By

Superclasses:  Expand list interview_age (Age Differentiation Test)
Age Differentiation Test
NIMH Data Archive
Common Data Element

Synonyms: Synonym

Existing Ids:

Preferred	CURIE	IRI
ILX:0115067	http://uri.interlex.org/base/ilx_0115067	
		http://uri.interlex.org/base/ilx_0100400

Log in to suggest term to community 

Back to search results 

Add new term

Export:   

Participant Age 
http://uri.interlex.org/base/ilx_0115067

PreferredId: ILX:0115067 Type: cde OWL Equivalent: owl:Class
Age at the time of study enrollment, expressed in number of years completed at the last birthday.

Version: 1

General Children Relationships Annotations Referenced By

Superclasses:  Expand list Participant Age
National Cancer Institute Common Data Element
Common data element
Standard specification
Narrative resource

Synonyms: Synonym

Existing Ids:

Preferred	CURIE	IRI
ILX:0115067	http://uri.interlex.org/base/ilx_0115067	

Log in to suggest term to community

Back to search results

Add new term

Common Data Elements

Home / Term Dashboard / ILX:0115067

Term View

Participant Age 
http://uri.interlex.org/base/ilx_0115067



 Suggest term to community

 Back to search results

 Add new term

PreferredId: ILX:0115067 Type: cde OWL Equivalent: owl:Class
Age at the time of study enrollment, expressed in number of years completed at the last birthday.

Export:   

Version: 1

General	Children	Relationships	Annotations	Referenced By														
<p>Superclasses:</p> <p> Expand list</p> <p>Participant Age </p> <p>National Cancer Institute Common Data Element </p> <p>Common data element </p> <p>Standard specification</p>	<p>Synonyms:</p> <table border="1"><thead><tr><th>Synonym</th><th>Type</th></tr></thead><tbody><tr><td>Age</td><td></td></tr><tr><td>Patient Age</td><td></td></tr><tr><td>Subject Age</td><td></td></tr></tbody></table>	Synonym	Type	Age		Patient Age		Subject Age		<p>Existing Ids:</p> <table border="1"><thead><tr><th>Preferred</th><th>CURIE</th><th>IRI</th></tr></thead><tbody><tr><td> ILX:0115067</td><td></td><td>http://uri.interlex.org/base/ilx_0115067</td></tr></tbody></table>	Preferred	CURIE	IRI	 ILX:0115067		http://uri.interlex.org/base/ilx_0115067		
Synonym	Type																	
Age																		
Patient Age																		
Subject Age																		
Preferred	CURIE	IRI																
 ILX:0115067		http://uri.interlex.org/base/ilx_0115067																
<p>OWL Equivalent: owl:ObjectProperty</p> <p> Show withdrawn relationships</p> <table border="1"><thead><tr><th>Relationship type</th><th>Term 2</th><th>Votes</th></tr></thead><tbody><tr><td>Records Instance Of</td><td>Age</td><td>+ 0   - 0</td></tr></tbody></table>					Relationship type	Term 2	Votes	Records Instance Of	Age	+ 0   - 0								
Relationship type	Term 2	Votes																
Records Instance Of	Age	+ 0   - 0																

Data Element Description in BIDS



Modality agnostic files

Brain Imaging Data Structure

The BIDS Specification ^

Introduction

Common principles

Modality agnostic files

Modality specific files ^

Longitudinal and multi-site studies

Extending the BIDS specification

Appendix ^

Changelog

The BIDS Starter Kit

Participants file

Template:

```
participants.tsv
participants.json
phenotype/<measurement_tool_name>.tsv
phenotype/<measurement_tool_name>.json
```

```
{
  "MeasurementToolMetadata": {
    "Description": "Adult ADHD Clinical Diagnostic Scale V1.2",
    "TermURL": "http://www.cognitiveatlas.org/task/id/trm_5586ff878155d"
  },
  "adhd_b": {
    "Description": "B. CHILDHOOD ONSET OF ADHD (PRIOR TO AGE 7)",
    "Levels": {
      "1": "YES",
      "2": "NO"
    }
  },
  "adhd_c_dx": {
    "Description": "As child met A, B, C, D, E and F diagnostic criteria",
    "Levels": {
      "1": "YES",
      "2": "NO"
    }
  }
}
```

Annotating Tabular Data - GUI

A	B	C	D	E	F
SITE_ID	SUB_ID	FILE_ID	DX_GROUP	DSM_IV_TR	AC
CALTECH	26042	Caltech_0051456	1	4	
CALTECH	51457	Caltech_0051457	1	4	
CALTECH	51458	Caltech_0051458	1	1	
CMU	50645	CMU_b_0050645	1	1	
CMU	50646	CMU_a_0050646	1	1	
CMU	50647	CMU_a_0050647	1	1	
KKI	50772	KKI_0050772	2	0	
KKI	50773	KKI_0050773	2	0	

PreferredId: NDA.CDE:2031448 Type: cde OWL Equivalent: owl:Class

Number assigned to site

Export:

Version: 5

General Children Relationships Annotations Referenced By

Superclasses:

Expand list

site_id (UC Provider Care)



UC Provider Care



NIMH Data Archive Common

Data Element

Synonyms:

Synonym

Type

Existing Ids:

Preferred CURIE

IRI

ILX:0310682

http://uri.interlex.org/base/ilx_0310682



NDA.CDE:2031448 <http://uri.interlex.org/NDA/uris/datadictionary/elements/2031448>

Site

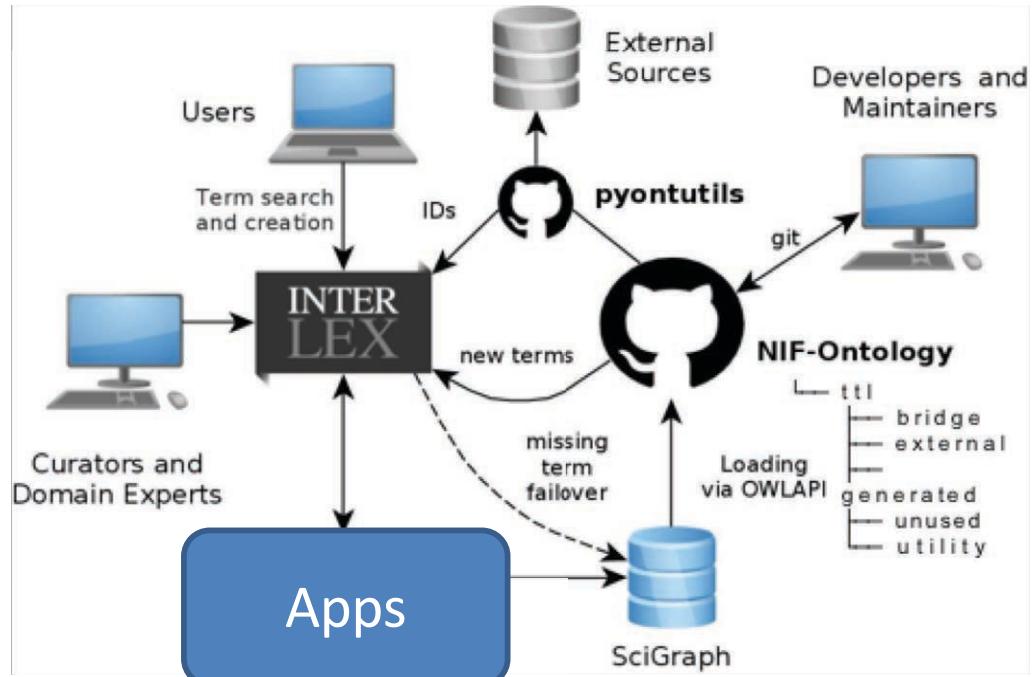
Preferred Id: ILX:0110649 Score: 29.061907 Type: term

Matches: existing_ids

InterLex

Community Managed Terminologies

- Dynamic lexicon of biomedical terms
- Constructed to help improve the way scientists communicate about their data
- Provide more powerful means of integrating data across distributed resources.
- InterLex allows for the association of data values (i.e. the value of a field or text within a field) to terminologies enabling the crowdsourcing of data-terminology mappings within and across communities.



Built on the foundation of NeuroLex

<https://interlex.org>

Term Search



Center for Reproducible Neuroimaging Computation

ABOUT

COMMUNITY RESOURCES

MORE RESOURCES

LITERATURE

MY ACCOUNT 3

Home / Term Dashboard / Term Search

Term Search

echo time



Search Term: echo time Result 1 through 100 of 304.

Search community terms only

First Previous **1** 2 3 4 Next Last

Echo Time

Preferred Id: NLX:149938 Score: 6.374187 Type: term

Time in ms between the middle of the excitation pulse and the peak of the echo produced (kx)

Matches: definition, label

Effective Echo Time

Preferred Id: NLX:149940 Score: 5.0728955 Type: term

The time in ms between the middle of the excitation pulse and the peak of the echo produced for kx

Matches: definition, label

TE

Preferred Id: BIRNLEX:2070 Score: 4.9995875 Type: term

Matches: synonyms, annotations

Duration

Preferred Id: PATO:0001309 Score: 2.1692846 Type: term

A process quality inhering in a bearer by virtue of the bearer's magnitude of the temporal extent be... [\[more\]](#)

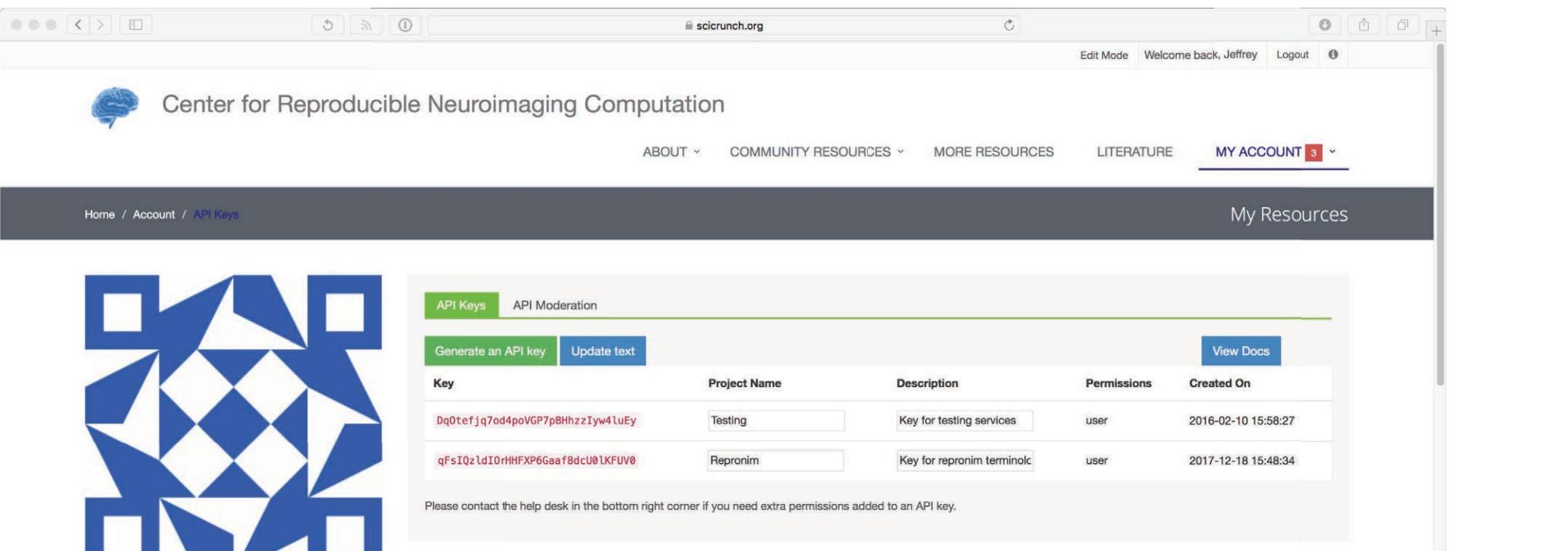
Matches: synonyms, annotations

Echo Number

Preferred Id: NLX:149935 Score: 1.9904175 Type: term

The echo number used in generating this image. In the case of segmented k-space, it is the effective... [\[more\]](#)

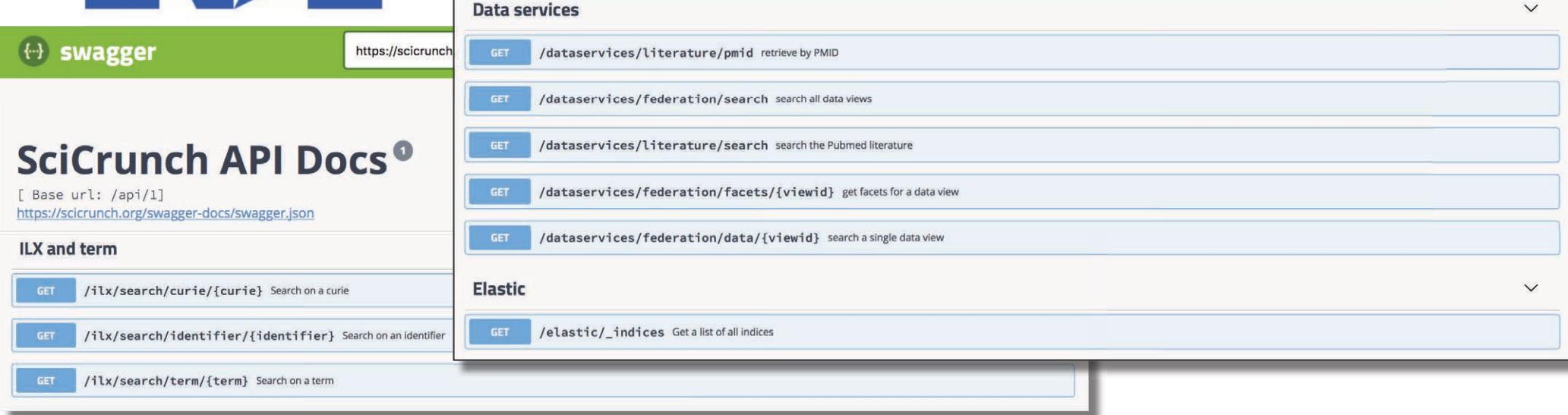
Access to Services



The screenshot shows the SciCrunch.org website interface. At the top, there's a navigation bar with links for 'ABOUT', 'COMMUNITY RESOURCES', 'MORE RESOURCES', 'LITERATURE', and 'MY ACCOUNT' (which has 3 notifications). Below this is a secondary navigation bar with 'Home / Account / API Keys' on the left and 'My Resources' on the right. The main content area features a large blue geometric logo on the left. In the center, there's a table titled 'API Keys' showing two entries:

Key	Project Name	Description	Permissions	Created On
Dq0tefjq7od4poVGP7pBHzzIyw4luEy	Testing	Key for testing services	user	2016-02-10 15:58:27
qFsIQzldIOrHHXP6Gaaaf8dcU0lKFUV0	Repronim	Key for repronim terminolo	user	2017-12-18 15:48:34

A note at the bottom of this section says: 'Please contact the help desk in the bottom right corner if you need extra permissions added to an API key.'



The screenshot shows the 'SciCrunch API Docs' page. It includes a 'swagger' button and a URL 'https://scicrunch.org/swagger-docs/swagger.json'. The main content area displays several API endpoints under sections like 'Data services' and 'Elastic'.

Data services

- GET /dataservices/literature/pmid retrieve by PMID
- GET /dataservices/federation/search search all data views
- GET /dataservices/literature/search search the Pubmed literature
- GET /dataservices/federation/facets/{viewid} get facets for a data view
- GET /dataservices/federation/data/{viewid} search a single data view

Elastic

- GET /elastic/_indices Get a list of all indices

Annotating Tabular Data - CLI

A	B	C	D	E	F	G	H	I	J	K	L
SITE_ID	SUB_ID	FILE_ID	DX_GROUP	DSM_IV_TR	AGE_AT_SCAN	SEX	HANDEDNESS_CATEGORY	HANDEDNESS_SCORES	FIQ	VIQ	PIQ
CALTECH	26042	Caltech_0051456		1	4	55.4	1 R		126	118	128
CALTECH	51457	Caltech_0051457		1	4	22.9	1 Ambi		107	119	93
CALTECH	51458	Caltech_0051458		1	1	39.2	1 R		93	80	108
CMU	50645	CMU_b_0050645		1	1	20	1 R		124	128	115
CMU	50646	CMU_a_0050646		1	1	21	1 R		108	100	115
CMU	50647	CMU_a_0050647		1	1	27	1 R		104	97	109
KKI	50772	KKI_0050772		2	0	12.77	1 R		100	98	
KKI	50773	KKI_0050773		2	0	10.84	1 R		63	124	
KKI	50774	KKI_0050774		2	0	10.64	1 R		56	124	
KKI	50775	KKI_0050775		2	0	11	1 R		85	101	

<https://github.com/incf-nidash/PyNIDM>

Search Term: SUB_ID

Search Results:

1: Narrow Interlex query

2: Change Interlex query string from: "SUB_ID"

3: Define my own term for this variable

Please select an option (1:3) from above: **2**

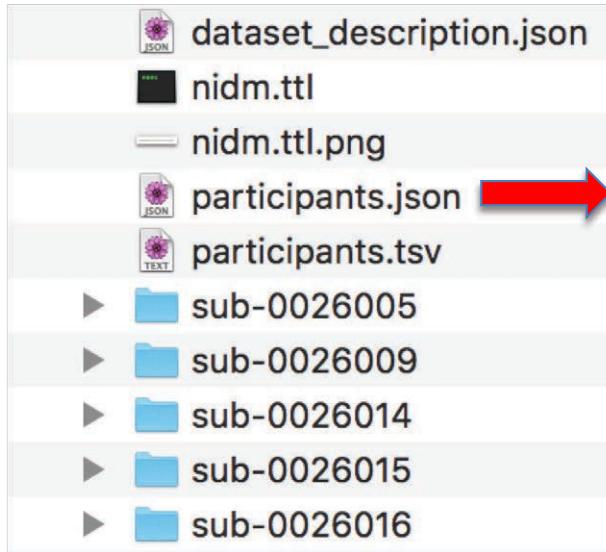
Please input new search term for CSV column: SUB_ID :*subject identifier*

Search Term: subject identifier

Search Results:

- | | |
|--|--|
| 1: Label: subjectkey (Consumer Satisfaction Survey) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred UI: |
| 2: Label: subjectkey (Continuous Performance Test) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 3: Label: subjectkey (Demographics and Screening) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 4: Label: subjectkey (Coping Self-Efficacy Scale) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 5: Label: subjectkey (Depression Survey) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 6: Label: subjectkey (Crovitz Handedness Scale) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 7: Label: subjectkey (Depression, Anxiety, and Stress Scale) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 8: Label: subjectkey (Current Menstrual State and Menstrual History) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 9: Label: subjectkey (Daily Stress Inventory) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 10: Label: subjectkey (Depressive Interpersonal Relationships Inventory) | Definition: The NDAR Global Unique Identifier (GUID) for research subject Preferred URL: http://u |
| 11: NIDM Constant: sio:Identifier URI: http://semanticscience.org/ontology/sio.owl#Identifier | |
| 12: Narrow Interlex query | |
| 13: Change Interlex query string from: "SUB_ID" | |
| 14: Define my own term for this variable | |

BIDS Annotations



```
{  
    "SITE_ID": {  
        "definition": "Number assigned to site",  
        "label": "site_id (UC Provider Care)",  
        "url": "http://uri.interlex.org/NDA/uris/datadictionary/elements/2031448"  
    },  
    "SUB_ID": {  
        "definition": "ndar:src_subject_id",  
        "label": "ndar:src_subject_id",  
        "url": "https://ndar.nih.gov/api/datadictionary/v2/dataelement/src_subject_id"  
    },  
    "FILE_ID": {  
        "definition": "nfo:filename",  
        "label": "nfo:filename",  
        "url": "http://www.semanticdesktop.org/ontologies/2007/03/22/nfo#filename"  
    },  
    "DX_GROUP": {  
        "definition": "ncit:Diagnosis",  
        "label": "ncit:Diagnosis",  
        "url": "http://ncitt.ncit.nih.gov/Diagnosis"  
    },  
    "DSM_IV_TR": {  
        "definition": "DSM-IV-TR PDD Category",  
        "label": "DSM-IV-TR",  
        "url": "https://github.com/dbkeator/nidm-local-terms/issues/1"  
    },  
    "AGE_AT_SCAN": {  
        "definition": "ncidb:Age",  
        "label": "ncidb:Age",  
        "url": "http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#Age"  
    },  
    "SEX": {  
        "definition": "ndar:gender",  
        "label": "ndar:gender",  
        "url": "https://ndar.nih.gov/api/datadictionary/v2/dataelement/gender"  
    },  
}
```

BrainVerse

- An **electronic laboratory notebook** built as a cross platform desktop application
- Enables users to **plan experiments, collect, analyze and reuse data, and collaborate**
- **Adds semantic annotation to data** with relevant metadata based on NeuroImaging data Model (NIDM) making experimental neuroimaging study more **reproducible**, and making data **FAIR**
- Addresses the limitations of existing Electronic Data Capture systems, such as REDCap and OpenClinica
 - Lack of semantic annotation, do not use underlying graph model, limited query capabilities, do not integrate with other desktop applications, integration and management of specific data types.

Session Number	Session Name	Task Name	Instrument Name	Status
1	Phone Screen	Call Subjects	terms-bu_demographics01-m32-BUDemographics	Not completed
2	Scan	Scanning	terms-abcd_ra01-m76-scan-list	Not completed
2	Scan	Measure Mindfulness	terms-camm01-m43-camm-edited	Not completed
3	GritScale	Grit Scale Measurement	terms-grit01-m74-12-gritscale	Not completed

<https://github.com/ReproNim/brainverse>

Section 3

Re-Usable Data and NIDM

Neuroimaging Data Model

Level 4

NIDM Dataset Descriptor

NIDM Access

Level 3

NIDM Experiment



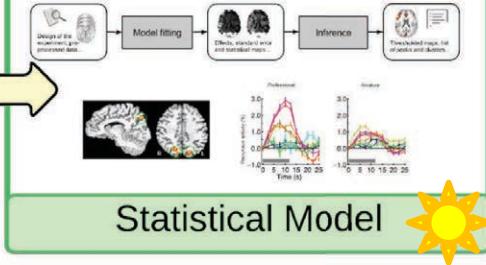
Data Acquisition

NIDM Workflow



Image Processing

NIDM Results



Statistical Model

and Query

Level 2

NIDM Core Vocabulary

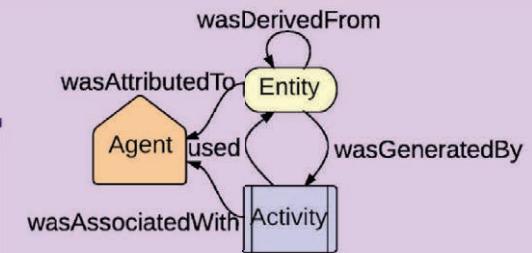
Level 1

PROV Family of Specifications

Level 0



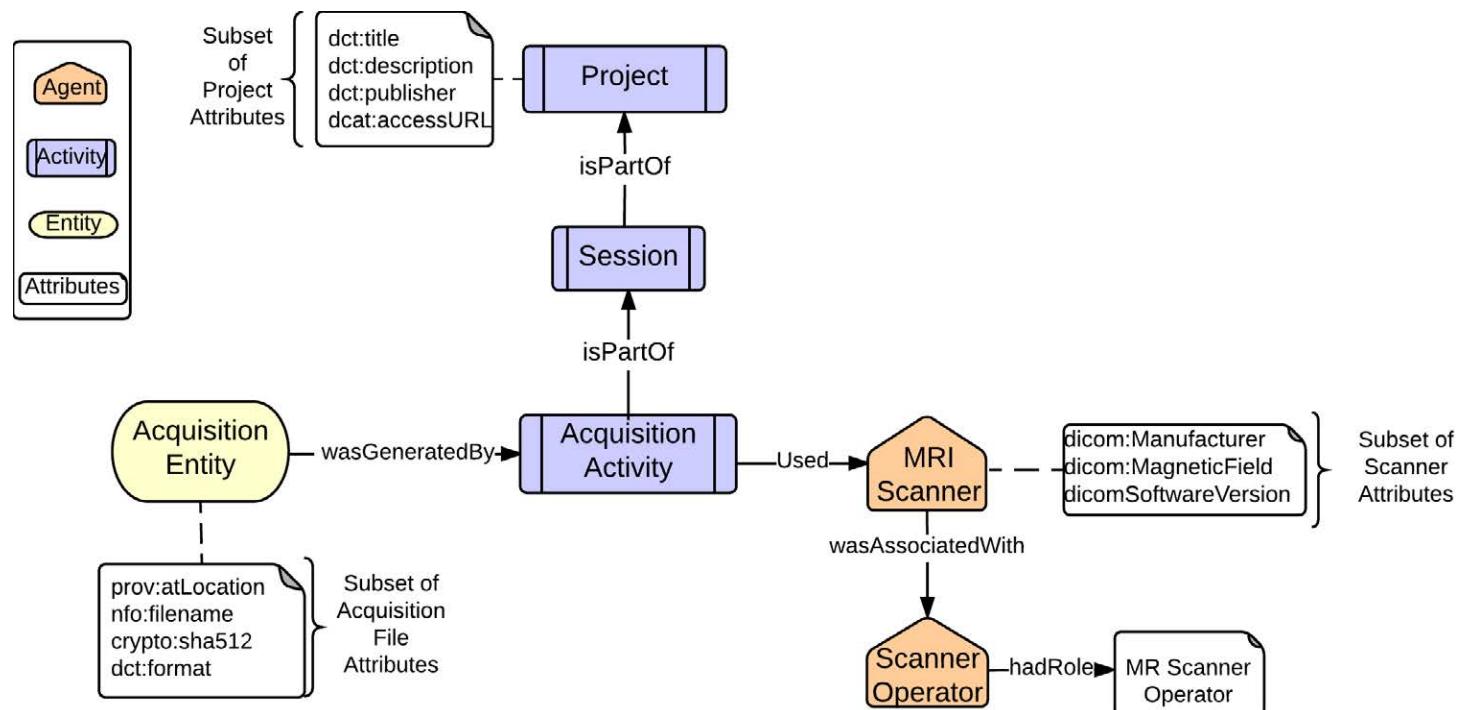
Semantic Web Technologies



<http://nidm.nidash.org/>

NIDM Experiment

- Representation of the experiment design, source data descriptions, and information on the participants and acquisition information.
- Metadata stored using simple Project->Session->Acquisition hierarchy
- Based on W3C Linked Data (www.w3.org/RDF/) and PROV (www.w3.org/TR/prov-overview/) models.
- Metadata described using subject-predicate-object tuples.
- Tuples composed of namespace:term where “namespace” is web-accessible ontology/terminology containing the precise, unambiguous definition of the “term”.



PyNIDM

[build](#) passing

A Python library to manipulate the [Neuro Imaging Data Model](#).

Dependencies

- [graphviz](#) (native package):
 - Fedora: `dnf install graphviz`
 - OS-X: `brew install graphviz`

creating a conda environment and installing the library (tested with OSX)

- `conda create -n pynidm_py3 python=3 pytest graphviz -y`
- `source activate pynidm_py3`
- `cd PyNIDM`
- `pip install -e .`
- you can try to run a test: `pytest`

NIDM Experiment Tools

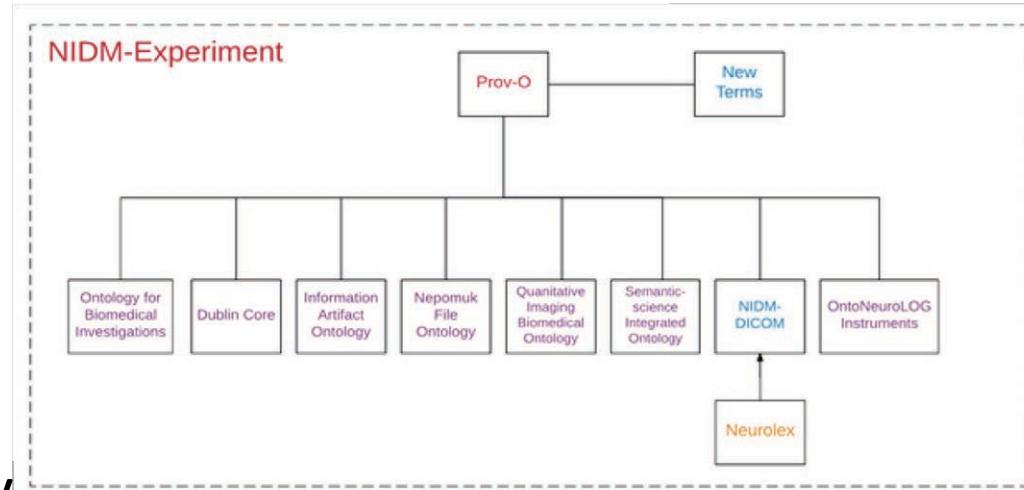
BIDSMRI2NIDM.py

- **Location:** PyNIDM/nidm/experiment/tools/BIDSMRI2NIDM.py
- **Description:** This tool will convert a BIDS MRI directory tree to a NIDM-Experiment document. Currently does not properly handle arbitrary Phenotype files. Will encode in NIDM document but namespace set to BIDS (<http://bids.neuroimaging.io/>) and term to simply variable name from phenotype file which won't de-reference....work in progress....

PyNIDM – API + Terminology

- Python-based API (github.com/incf-nidash/PyNIDM) following the simple organizational structure of NIDM-Experiment documents.

- Functions to create, query, export, import, and transform NIDM-Experiment documents.



- NIDM Terminology / Ontology
 - NIDM ontology reuses existing terms from several other where appropriate definitions exist.
 - NIDM-Experiment ontology: github.com/incf-nidash/nidm/tree/master/nidm/nidm-experiment/terms
 - DICOM-tag ontology: github.com/incf-nidash/dicom-ontology

PyNIDM Terminology Alignment

```
#BIDS Participants file -> NIDM constants      NIDM SUBJECTID =
mappingsparticipants = { "participant_id": QualifiedName(provNamespace("ndar",NDAR),
Constants.NIDM SUBJECTID, "sex": "src_subject_id")
Constants.NIDM_GENDER, "age": nidm:_88f65012-e44b-11e7-b9e2-6c4008b8f03e a
Constants.NIDM AGE, "gender": prov:Agent,
Constants.NIDM_GENDER, "diagnosis": prov:Person ;
Constants.NIDM_DIAGNOSIS}                                ndar:src_subject_id "sub-10159"^^xsd:string .
```



```
DICOM =
Namespace("http://neurolex.org/wiki
/Category:DICOM_term/")

nidm:_88f8dcec-e44b-11e7-8ba0-
6c4008b8f03e a sio:file,
    prov:Entity ;
    dicom:EchoTime 3.31e-03 ;

json_keys = { ##Image terms "run" :
    Constants.NIDM_ACQUISITION_ENTITY, "ImageType" :
    Constants.DICOM["ImageType"],
    "ManufacturerModelName" :
    Constants.DICOM["ScanningSequence"],
    "SequenceVariant" : Constants.DICOM["SequenceVariant"],
    "ScanOptions" : Constants.DICOM["ScanOptions"],
    "MRAcquisitionType" :
    Constants.DICOM["MRAcquisitionType"],
    "SequenceName" : Constants.DICOM["SequenceName"],
    "RepetitionTime" : Constants.DICOM["RepetitionTime"],
    "EchoTime" : Constants.DICOM["EchoTime"],
```

What is Linked Data?

“The term Linked Data is used to describe a *method* of exposing, sharing, and connecting data via dereferenceable URIs on the Web.” – Wikipedia

Linked Data is a protocol component of the Semantic Web, using URIs, Real World Objects, Content Negotiation and RDF documents to share information on the web.



The Semantic Web

- “The semantic web is a vision of information that is understandable by computers, so computers can perform more of the tedious work involved in finding, combining, and acting upon information on the web.” – Wikipedia
- “I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers.”
 - Tim Berners-Lee

Linked Data Principles

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up (dereference) those names.
3. When someone looks up a URI, provide useful information.
4. Include links to other URIs so that they can discover more things.

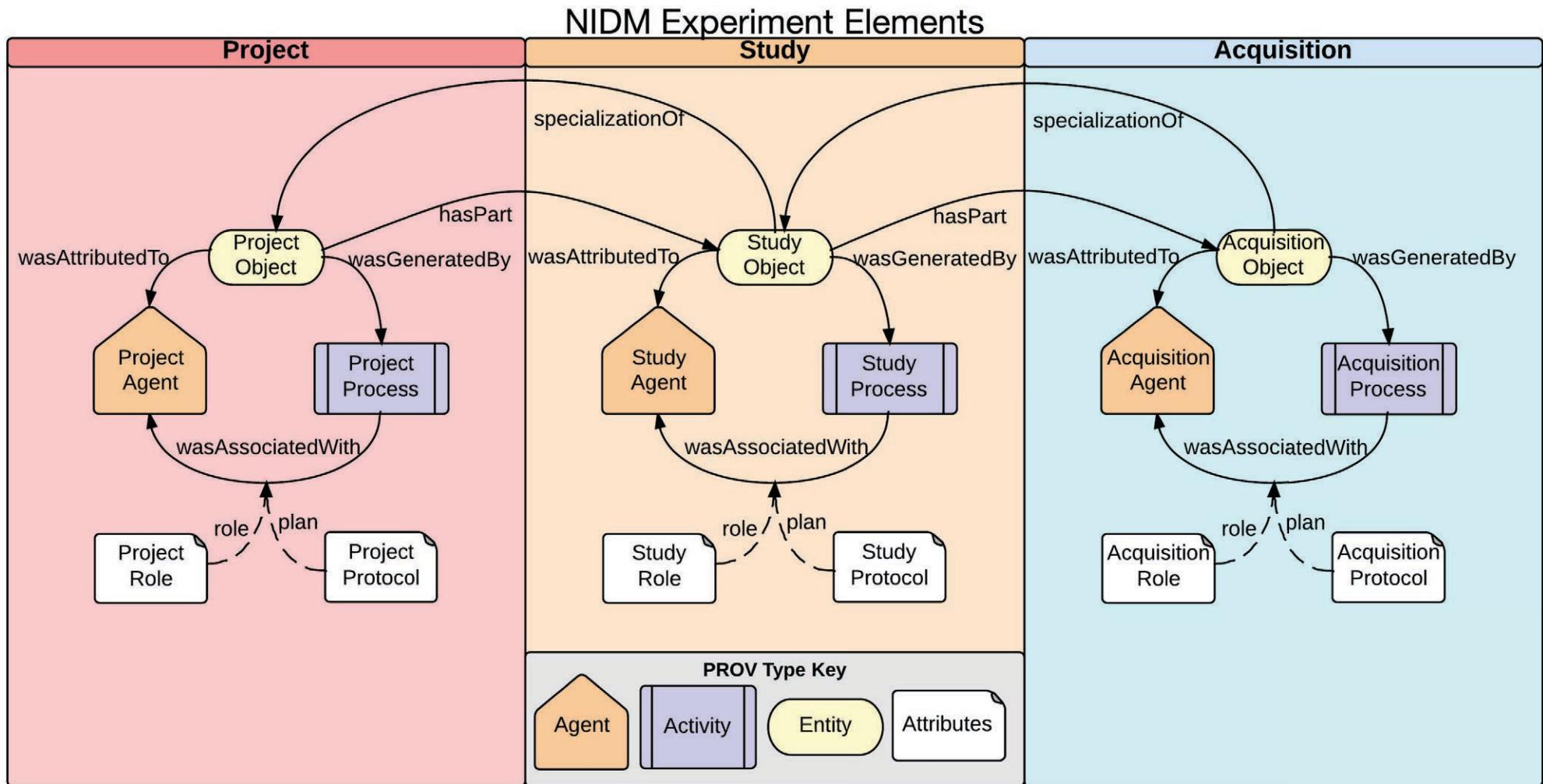
“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data. Like the web of hypertext, the web of data is constructed with documents on the web.

However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF,. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow.”

Resource Description Framework (RDF)

- A data model
 - A way to model data
 - i.e. Relational databases use relational data model
- RDF is a triple data model
- Labeled Graph
- Subject, Predicate, Object
- <Cerebellum> <is part of> <Hind Brain>

NIDM Graph Model



NIDM Resources

NIDM Overview

<http://nidm.nidash.org/specs/nidm-overview.html>

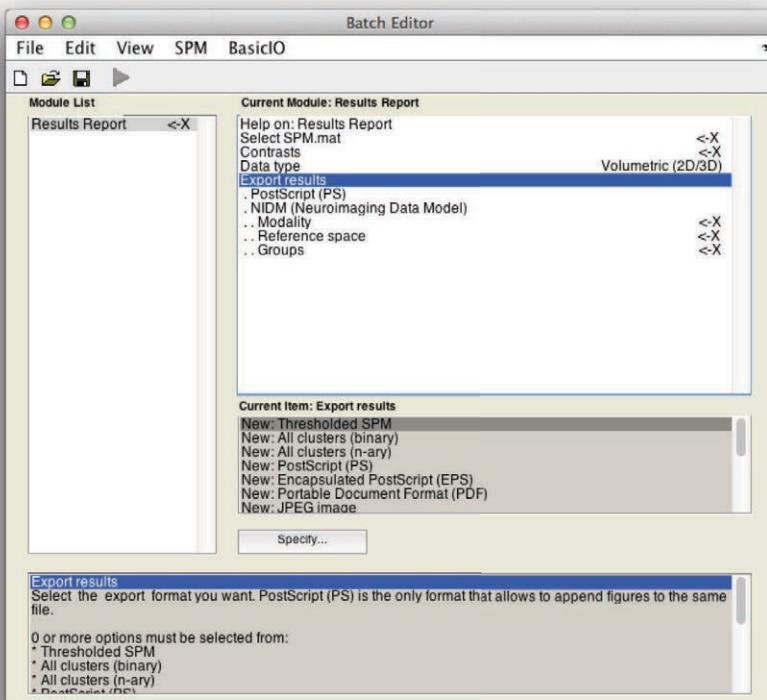
NIDM Primer

<http://nidm.nidash.org/specs/nidm-primer.html>

Getting Started with NIDM-Results

<http://nidm.nidash.org/getting-started/>

NIDM Results



SCIENTIFIC DATA

Altmetric: 51 Citations: 3 More detail >

Article | OPEN

Sharing brain mapping statistical results with the neuroimaging data model

Camille Maumet , Tibor Auer, Alexander Bowring, Gang Chen, Samir Das, Guillaume Flandin, Satrajit Ghosh, Tristan Glatard, Krzysztof J. Gorgolewski, Karl G. Helmer, Mark Jenkinson, David B. Keator, B. Nolan Nichols, Jean-Baptiste Poline, Richard Reynolds, Vanessa Sochat, Jessica Turner & Thomas E. Nichols

Scientific Data 3, Article number: 160102 (2016) doi:10.1038/sdata.2016.102

Received: 21 June 2016
Accepted: 21 September 2016
Published online: 06 December 2016

Download Citation

Medical research Research data

incf-nidash / nidmresults-fsl

Code Issues 13 Pull requests 1 Projects 1 Wiki Insights

Watch 14 Star 1 Fork 9

A python library to export FSL's feat results to NIDM-Results <http://nidm.nidash.org/specs/nidm-res...>



PyNIDM: Neuroimaging Data Model in Python

A Python library to manipulate the [Neuroimaging Data Model](<http://nidm.nidash.org>).

build passing

Contents

- 1 Dependencies
- 2 Creating a conda environment and installing the library (tested with OSX)
 - 2.1 macOS
- 3 NIDM-Experiment Tools
 - 3.1 Query
 - 3.2 BIDS MRI Conversion to NIDM
 - 3.3 CSV File to NIDM Conversion

3.2 BIDS MRI Conversion to NIDM

This program will convert a BIDS MRI dataset to a NIDM-Experiment RDF document. It will parse phenotype information and simply store variables/values and link to the associated json data dictionary file.

While we're migrating to using 'click', this tools is still buried in the tools directory of the repo

```
$ ./nidm/experiment/tools/BIDSMRI2NIDM.py -d [ROOT BIDS DIRECT] -bidsignore
```

NIDM.TTL

```
ncicb:Age "0.556"^^xsd:string ;
prov:wasGeneratedBy nidm:_61ee85c6-717f-11e8-b46b-0800277be888 .

nidm:_61ee85cb-717f-11e8-b46b-0800277be888 a onli:assessment-instrument,
      nidm:AcquisitionObject,
      prov:Entity ;
bids:ctopp.pa.comp "1.198531065"^^xsd:string ;
bids:ctopp.rln.raw "-1.104109688"^^xsd:string ;
bids:mcdecode "1.514966059"^^xsd:string ;
bids:mcpassfluency "1.680192391"^^xsd:string ;
bids:mrt "1.358940694"^^xsd:string ;
bids:piat.l.raw "0.766663412"^^xsd:string ;
bids:piat.r.raw "1.07060038"^^xsd:string ;
bids:ppvt.raw "1.303491936"^^xsd:string ;
bids:sspan2.raw "1.876107955"^^xsd:string ;
bids:wasi.matr.raw "1.413554228"^^xsd:string ;
ncicb:Age "1.416"^^xsd:string ;
prov:wasGeneratedBy nidm:_61ee85ca-717f-11e8-b46b-0800277be888 .

nidm:_61ee85ce-717f-11e8-b46b-0800277be888 a nidm:AcquisitionObject,
      prov:Entity ;
nidm:AcquisitionModality nidm:MagneticResonanceImaging ;
nidm:hasImageContrastType nidm:T1Weighted ;
nidm:hasImageUsageType nidm:Anatomical ;
nfo:filename "/sub-01/anat/sub-01_T1w.nii.gz"^^xsd:string ;
prov:wasGeneratedBy nidm:_61ee85cd-717f-11e8-b46b-0800277be888 .

nidm:_61ee85d0-717f-11e8-b46b-0800277be888 a nidm:AcquisitionObject,
      prov:Entity ;
nidm:AcquisitionModality nidm:MagneticResonanceImaging ;
nidm:hasImageContrastType nidm:T1Weighted ;
nidm:hasImageUsageType nidm:Anatomical ;
```

RDFLIB SPARQL Query

```
import rdflib

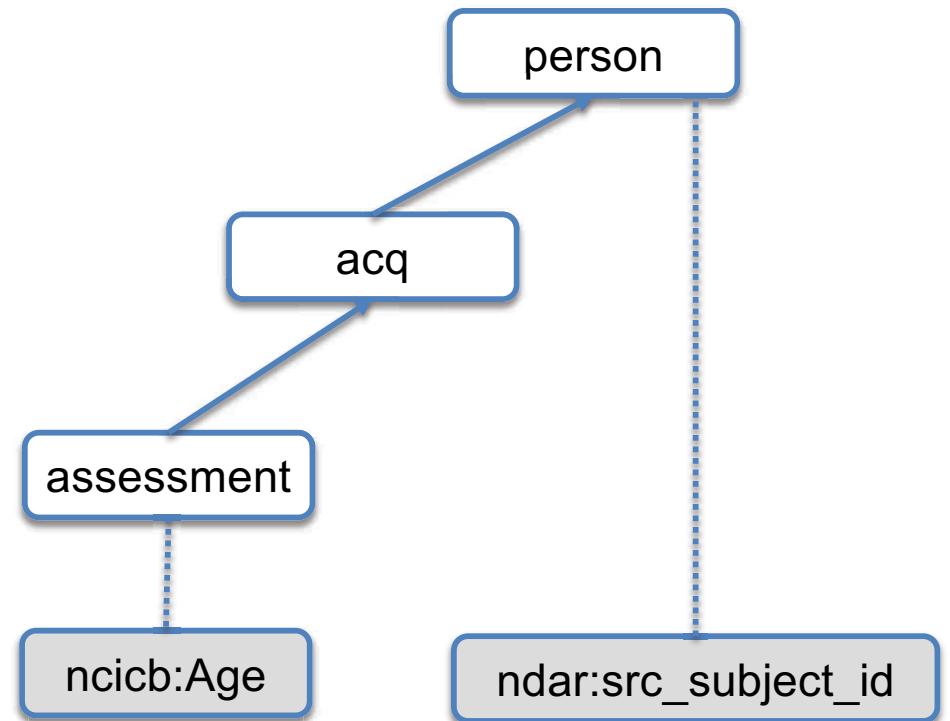
from argparse import ArgumentParser

parser = ArgumentParser()
# parse command line arguments
parser.add_argument('-nidm', dest='nidm_file', required=True, help="NIDM-Exp RDF File to import")
args = parser.parse_args()

g=rdflib.Graph()
g.parse(args.nidm_file, format='ttl')

qres = g.query(
    """SELECT DISTINCT ?id ?age ?assessment
    WHERE {
        ?assessment prov:wasGeneratedBy ?acq .
        ?acq prov:wasAssociatedWith ?person .
        ?assessment ncicb:Age ?age .
        ?person ndar:src_subject_id ?id
    }"""
)

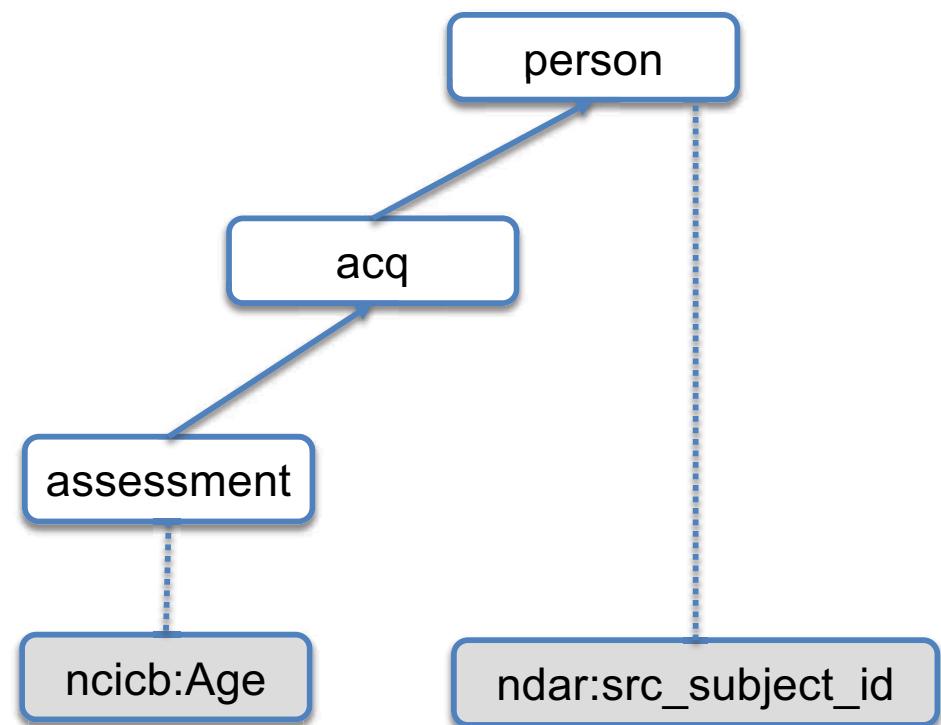
for row in qres:
    print("%s - %s - %s" % row)
```



RDFLIB SPARQL Query

```
(section1) vagrant@nitrcce:~/nidm-training$ python rdf-age-query.py -nidm  
~/section1/Repro_BIDS/nidm.ttl  
  
sub-02 - 24 - http://purl.org/nidash/nidm#_d6c8b9f1-debb-11e8-b0cd-0800277be888  
sub-01 - 32 - http://purl.org/nidash/nidm#_d6c8b9ed-debb-11e8-b0cd-0800277be888
```

```
g=rdflib.Graph()  
g.parse(args.nidm_file, format='ttl')  
  
qres = g.query(  
    """SELECT DISTINCT ?id ?age ?assessment  
    WHERE {  
        ?assessment prov:wasGeneratedBy ?acq .  
        ?acq prov:wasAssociatedWith ?person .  
        ?assessment ncicb:Age ?age .  
        ?person ndar:src_subject_id ?id  
    }"""")  
  
for row in qres:  
    print("%s - %s - %s" % row)
```



NIDM Resources

- Main Website: <http://nidm.nidash.org/>
- Mailing List: incf-nidash-nidm@googlegroups.com
- Google Drive: <http://bit.ly/NIDMDrive>
- GitHub
 - NIDM Specs Repository: <https://github.com/INCF-NIDASH/nidm-specs>
 - NIDM-Results API: <https://github.com/incf-nidash/nidmresults>
 - SPM Software API: <https://github.com/incf-nidash/nidmresults-spm>
 - FSL Software API: <https://github.com/incf-nidash/nidmresults-fsl>
 - PyNIDM API Repository: <https://github.com/incf-nidash/PyNIDM>
 - NIDM Training Materials (a bit older): <https://github.com/incf-nidash/nidm-training>
 - Other interesting repos: <https://github.com/incf-nidash>

