Brain imaging datasharing

The development of NIDM

The missing principles and tools

Jean-Baptiste Poline

UC Berkeley Brain Imaging Center

INCF Nidash TF

incf

Outline

- Background / Introduction : Software development and sociological aspects
- Why NIDM ? How is it developed ?
- Current use cases
  - Experimental data
  - Freesurfer
  - FMRI statistical results (SPM, FSL)
- Future of NIDM
  - Link with other projects
  - The big picture : standardizing the meta data

incf

Software development in neuroimaging: background

- Groups want an easy media for dissemination of their methods (SPM, FSL, Afni, etc)
- Code is often open, but often a “lab” enterprise
- Generally, software is still poorly considered in research (e.g. numpy author)
- Poor interoperability between tools: No W3C for neuroimaging. Nifti experience.
- Publications (the currency) are not adapted

No provenance of results and data, little testing

incf

Sociological consequences

- Groups have to “promote” their software
  - A medium for competitive advantage
  - Software suite silos

- Code to reproduce paper is not available and is not reviewed – grants panels do not often care
- Poor code development standard & training of scientists

- Out of NIH 400M USD for fMRI – how much for software development ?
- How much of this across institutions/countries ?

incf

Consequences for the scientific community

- Lack of harmonization and standard development leads to **inefficient** research (eg Nifti standard)
- Lack of re-usability: A big **waste of resources**
- **Lack of** detailed **provenance** information
- Lack of good testing and development standard: **reproducibility issue (**Donoho/Claerbout)

We need to respond to the reproducibility crisis

incf

Software: The missing principles

- **Publicly funded** research should lead to public research products. Not only help me get the next grant.
- Research is about **advancing knowledge:** Software should help to:
  - Reproduce and validate - Tested
  - Provide with provenance
  - Help us be efficient (limited resources)
- Adopting a culture of software and data sharing: **long term versus short term**
  - Re-use rather than re-develop
  - Collaborate when ever it is possible

incf

NIDASH-Neuroimaging Data Model

Goals

- Comprehensive data sharing
- Enhance reproducibility
- Enhance re-usability
- Increase interoperability and efficiency
  - Discover data
  - Access and use data
- Enable new research and idea

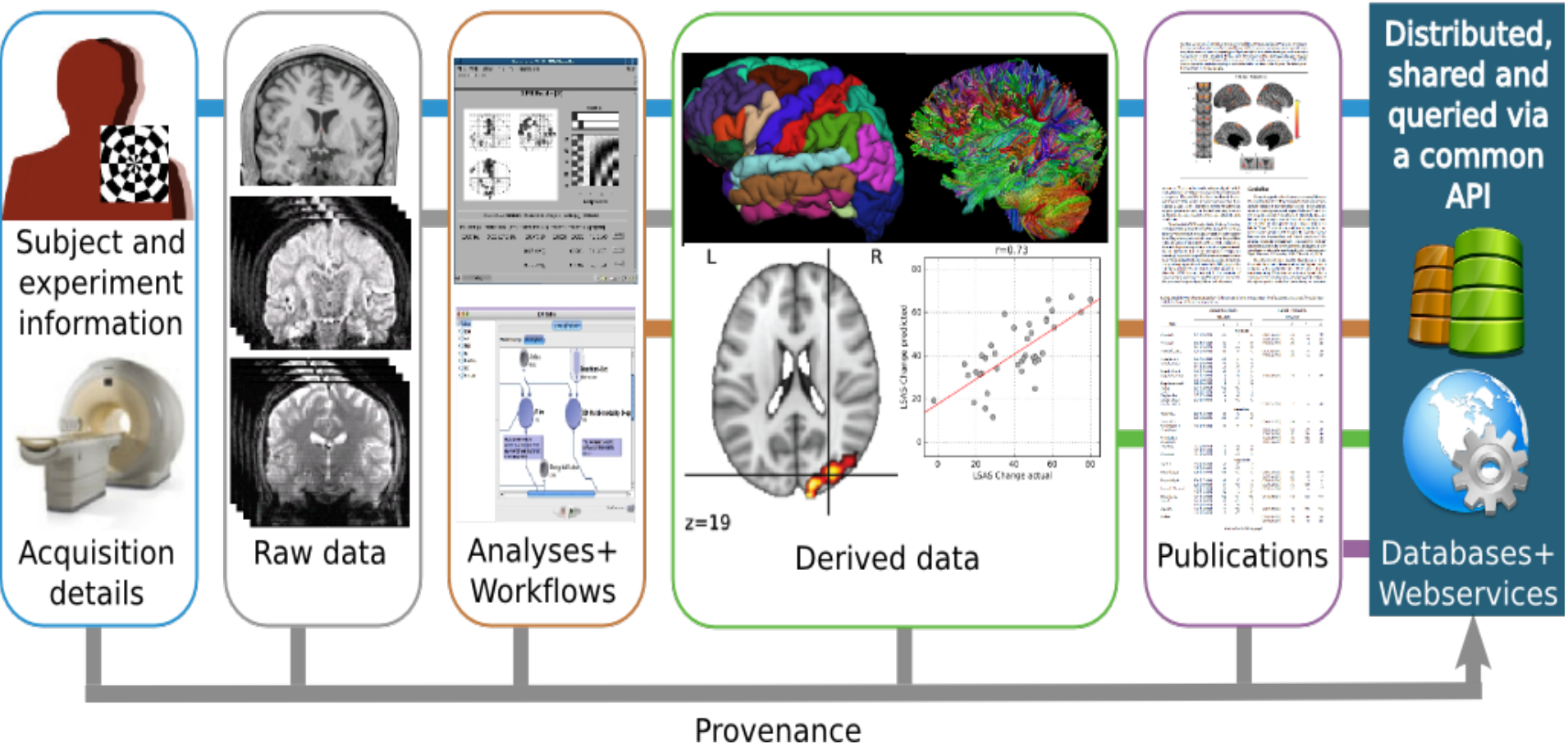
Challenges

- No easy tools
- Missing metadata
- Undiscoverable
- Multiple software
- Limited provenance
- Very limited funding
- **No common standards**

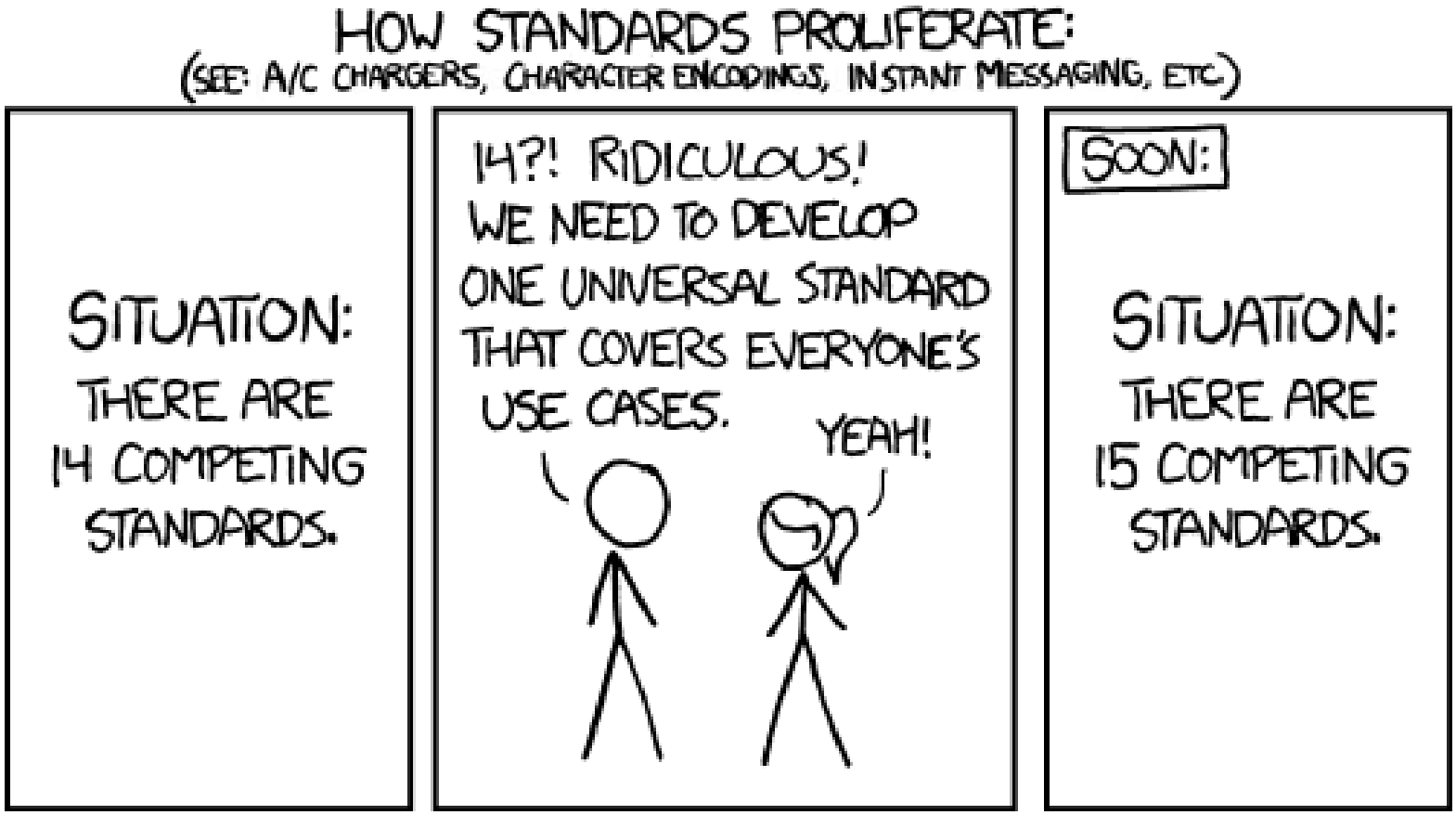
We need a common language, and a method to construct it

incf

NIDM: What are we trying to do ?



How do we avoid this ?



NIDM : solving for both technical issues and social engineering

- NIDM: A data model shared and co-developed
- NIDM development : Nidash methodology
  - Weekly call and crowd sourced minutes (google documents)
  - Git / Github for the development, BSD license
  - Pull requests / issues for discussion on terms
  - Hackathon
- Tools developers on board
- Technical: semantic web solution

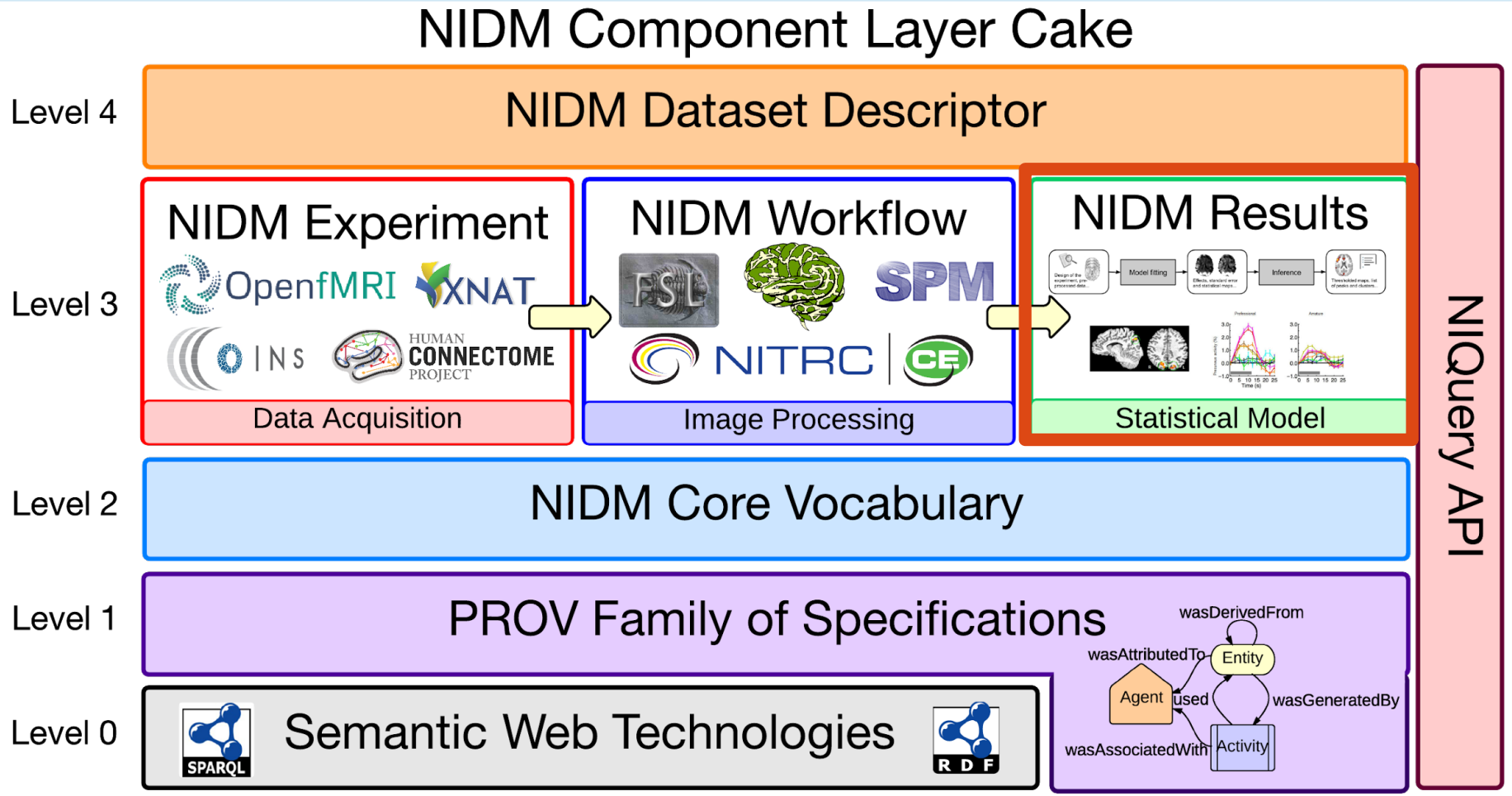
incf


G.F.

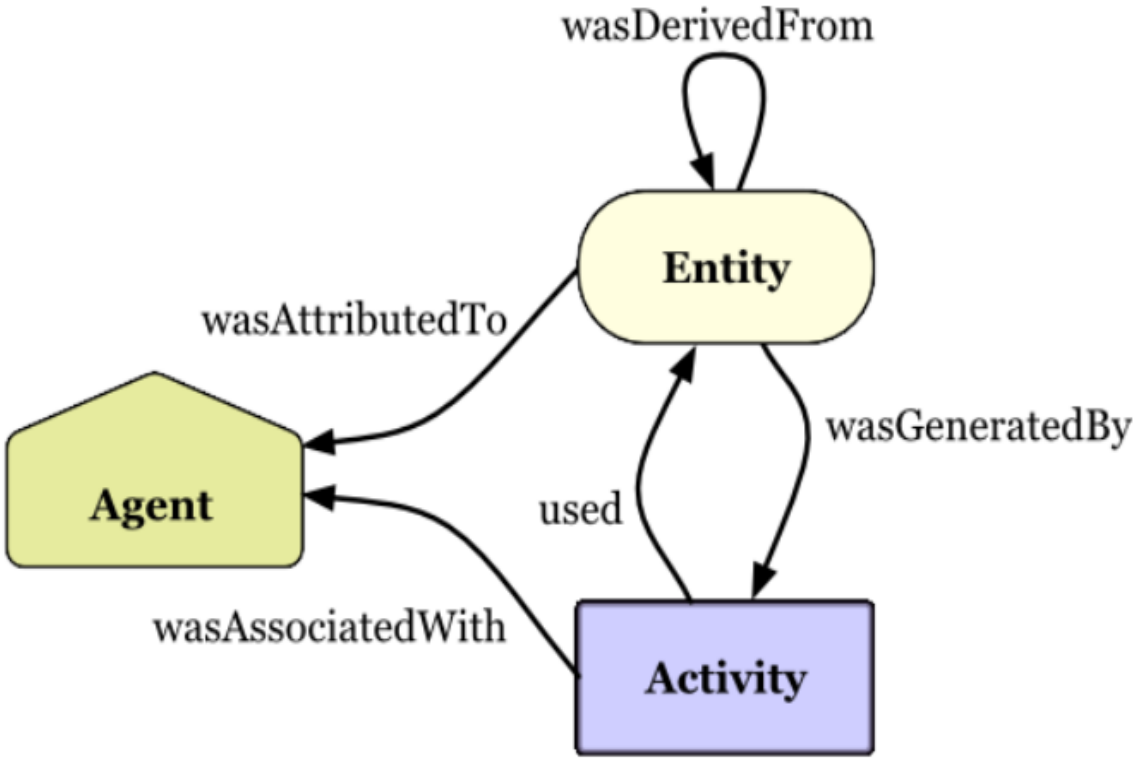
Community development – get the right people on board

brainhack.org

So, what is NIDM based on ?

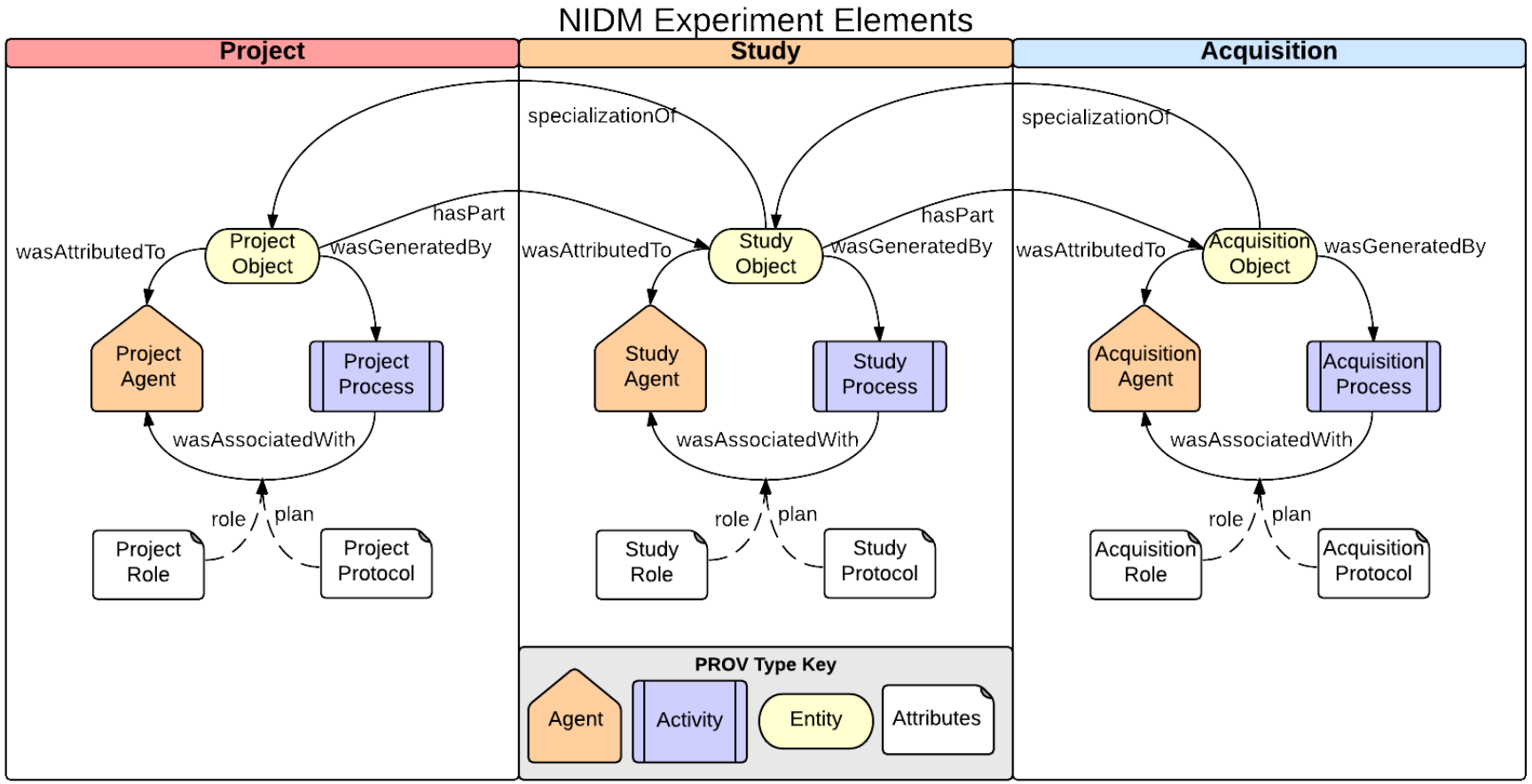


PROV Model



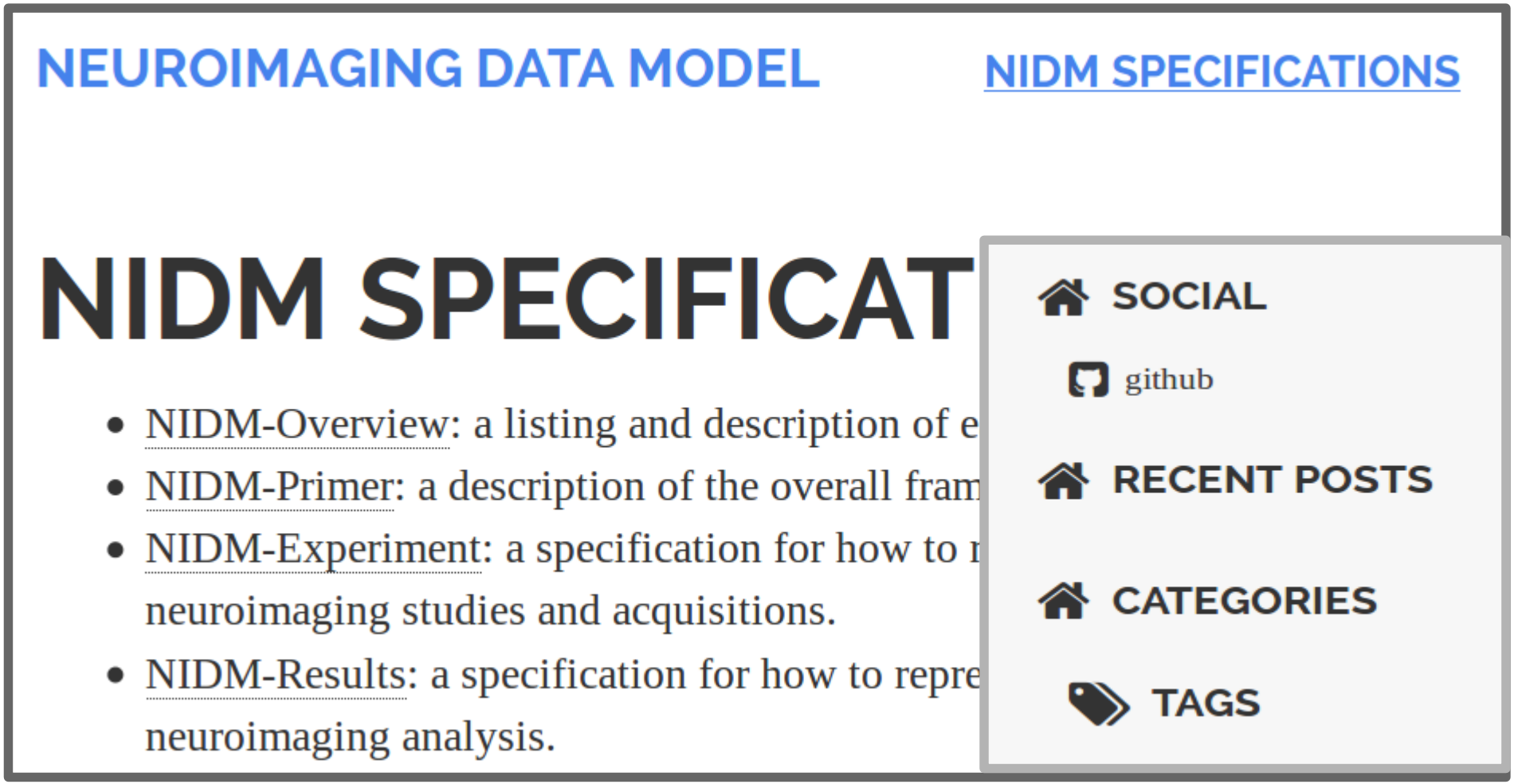
http://www.w3.org/TR/prov-primer/

NIDM-Experiment



Credits: Nolan Nichols et al.

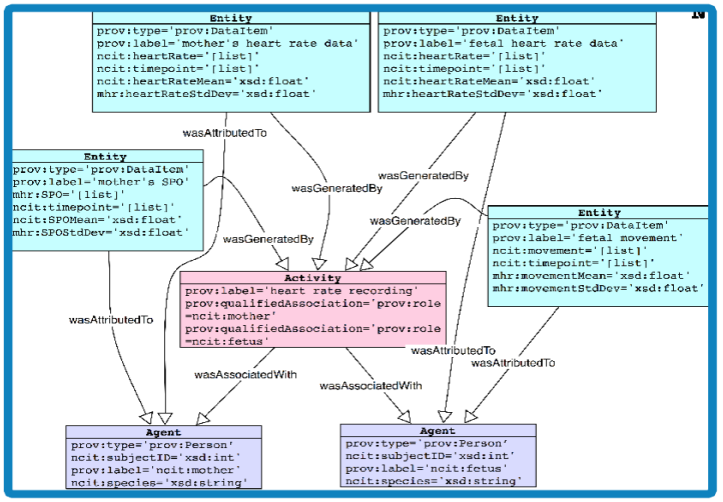
How do we involve larger community



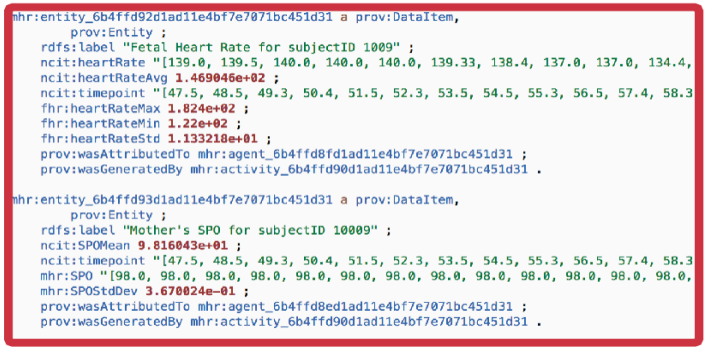
Use Case 1: Conte DB

Mother	Fetus	Gestation and Task	TimePoint(sec)	MHR	FHR	FMI	SPO2
10009	1009	T3 REST	47.5	86	139	0	98
10009	1009	T3 REST	48.5	87	139.5	31	98
10009	1009	T3 REST	49.3	87	140	0	98
10009	1009	T3 REST	50.4	88	140	31	98
10009	1009	T3 REST	51.5	91	140	0	98
10009	1009	T3 REST	52.3	91	139.33	0	98
10009	1009	T3 REST	53.5	92	138.4	31	98
10009	1009	T3 REST	54.5	93	137	31	98
10009	1009	T3 REST	55.3	93	137	0	98
10009	1009	T3 REST	56.5	92	134.4	31	98
10009	1009	T3 REST	57.4	91	132.25	0	98
10009	1009	T3 REST	58.3	91	133.33	0	98
10009	1009	T3 REST	59.5	90	134.8	31	98
10009	1009	T3 REST	60.5	88	136.25	31	98
10009	1009	T3 REST	61.3	88	138	0	98
10009	1009	T3 REST	62.2	87	138	0	98
10009	1009	T3 REST	63.5	86	138	31	98
10009	1009	T3 REST	64.3	86	137.33	0	98
10009	1009	T3 REST	65.2	87	136.75	31	98
10009	1009	T3 REST	66.5	88	136	31	98
10009	1009	T3 REST	67.3	88	137	0	98
10009	1009	T3 REST	68.2	91	136.25	0	98
10009	1009	T3 REST	69.4	94	134	31	98
10009	1009	T3 REST	70.4	94	131.67	0	98
10009	1009	T3 REST	71.3	94	130	31	98
10009	1009	T3 REST	72.5	93	129.8	31	98
10009	1009	T3 REST	73.2	93	129	0	98
10009	1009	T3 REST	74.3	92	129	0	98
10009	1009	T3 REST	75.5	91	129	31	98
10009	1009	T3 REST	76.5	91	129.75	31	98
10009	1009	T3 REST	77.3	91	130.67	0	98
10009	1009	T3 REST	78.5	90	133.4	31	98
10009	1009	T3 REST	79.4	90	137	0	98
10009	1009	T3 REST	80.2	88	138	0	98

NIDM Model



Encode



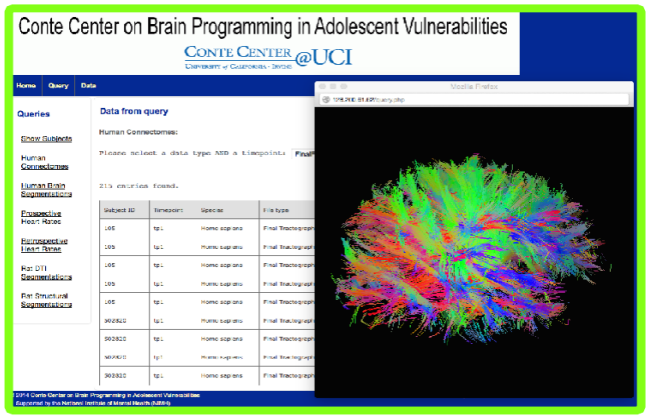
```
mhr:entity_6b4ffd92d1ad11e4bf7e7071bc451d31 a prov:DataItem,
        prov:Entity ;
    rdfs:label "Fetal Heart Rate for subjectID 1009" ;
    ncit:heartRate "[139.0, 139.5, 140.0, 140.0, 140.0, 139.33, 138.4, 137.0, 137.0, 134.4,
    ncit:heartRateAvg 1.469046e+02 ;
    ncit:timepoint "[47.5, 48.5, 49.3, 50.4, 51.5, 52.3, 53.5, 54.5, 55.3, 56.5, 57.4, 58.3
    fhr:heartRateMax 1.824e+02 ;
    fhr:heartRateMin 1.22e+02 ;
    fhr:heartRateStd 1.133218e+01 ;
    prov:wasAttributedTo mhr:agent_6b4ffd8fd1ad11e4bf7e7071bc451d31 ;
    prov:wasGeneratedBy mhr:activity_6b4ffd90d1ad11e4bf7e7071bc451d31 .

mhr:entity_6b4ffd93d1ad11e4bf7e7071bc451d31 a prov:DataItem,
        prov:Entity ;
    rdfs:label "Mother's SPO for subjectID 10009" ;
    ncit:SPOMean 9.816043e+01 ;
    ncit:timepoint "[47.5, 48.5, 49.3, 50.4, 51.5, 52.3, 53.5, 54.5, 55.3, 56.5, 57.4, 58.3
    mhr:SPO "[98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0, 98.0,
    mhr:SPOStdDev 3.670024e-01 ;
    prov:wasAttributedTo mhr:agent_6b4ffd8ed1ad11e4bf7e7071bc451d31 ;
    prov:wasGeneratedBy mhr:activity_6b4ffd90d1ad11e4bf7e7071bc451d31 .
```

Store

Virtuoso Database

Retreive

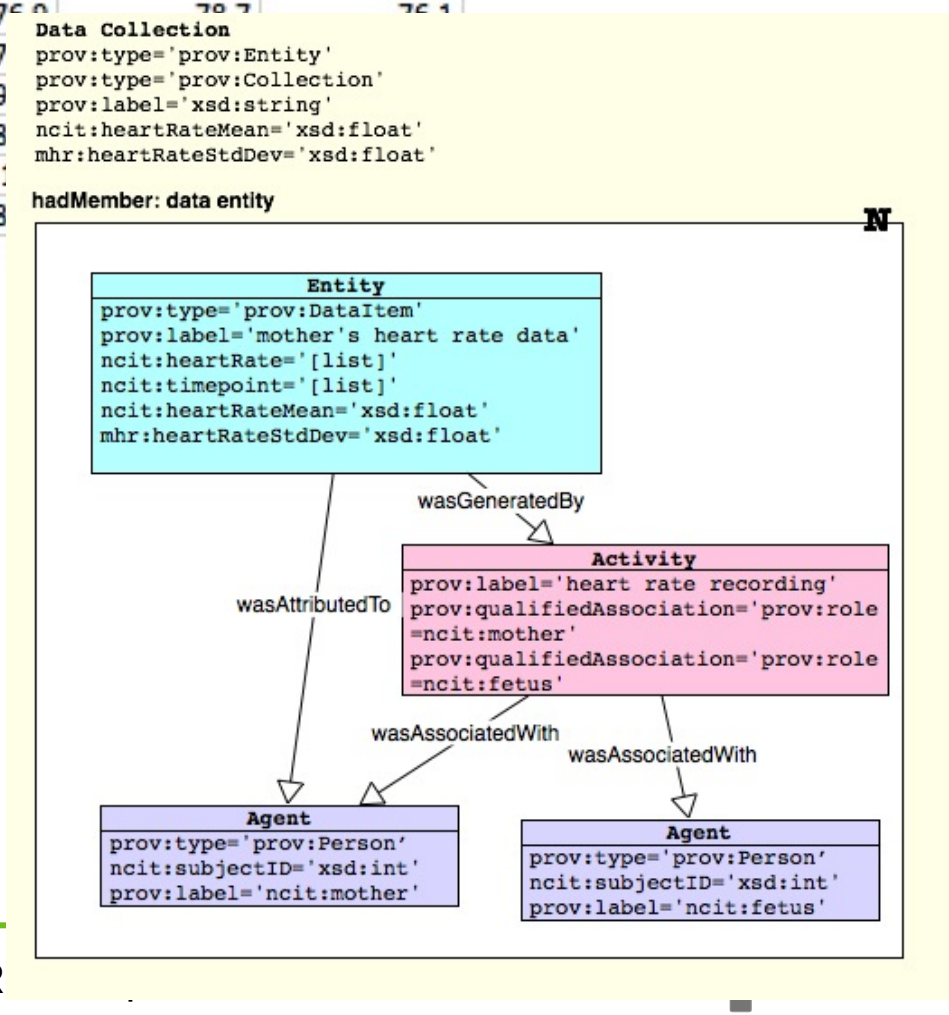


Nichols B.N., Keator D.B., et. al. Application of the Neuroimaging Data Model to Represent and Exchange Primary and Derived Data. Human Brain Mapping, Honolulu, Hawaii. 2015.

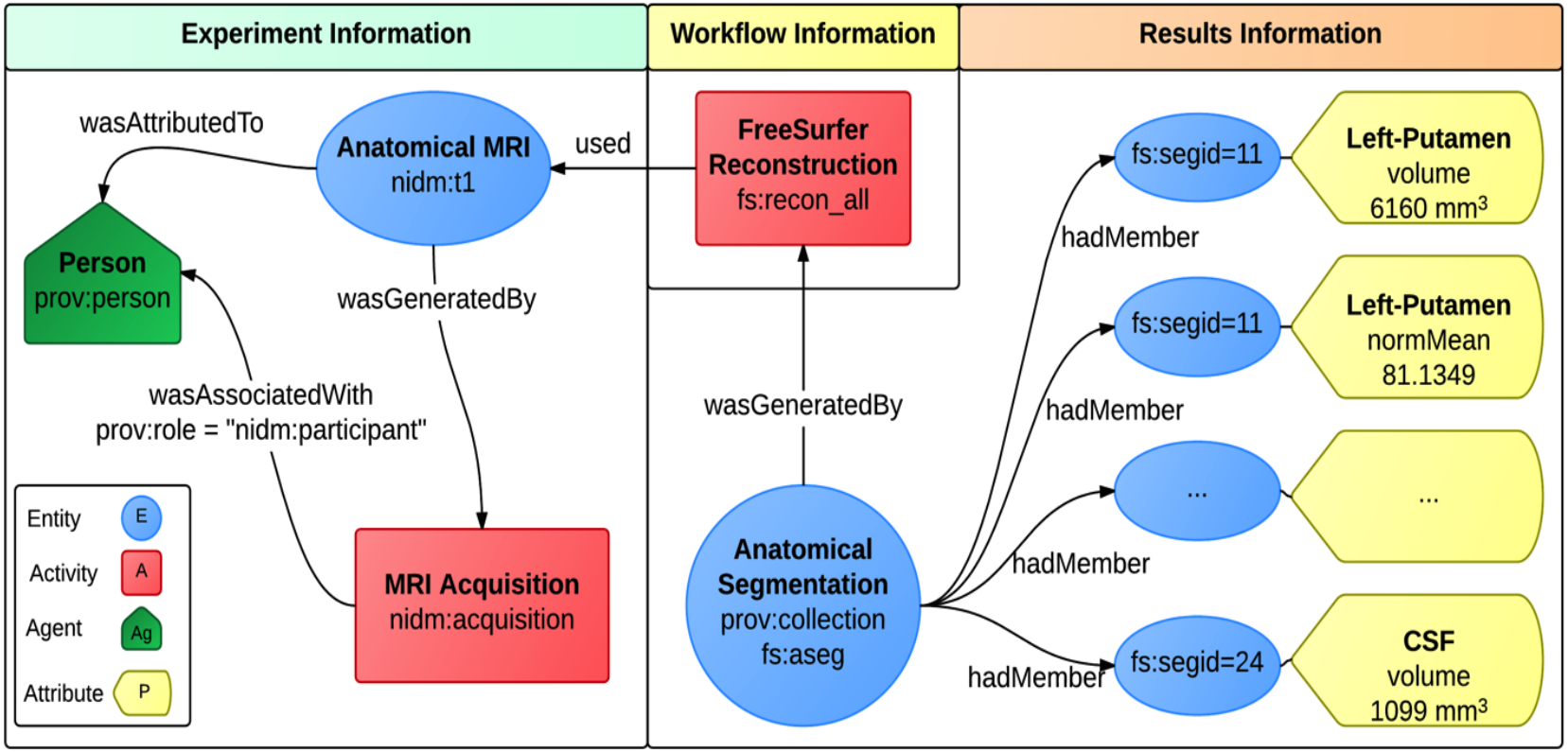
Heart rate description

A	B	C	D	E	F	G	H
subid	nsubid	A00.mhr.v180	A00.mhr.v179	A00.mhr.v178	A00.mhr.v177	A00.mhr.v176	A00.mhr.v175
30283	561	86.4	88.4	83.1	73.7	78.5	79.1
30378	590	75.9	74.7	78.6	75.4	72.7	78.1
30407	603	74.8	75.8	75.2	7		
30426	612	75.8	74.7	73.2	7		
30446	608	92.2	92.9	93.2	9		
30455	622	87.1	84.6	82.3	8		
30461	630	99.6	100.1	99.8			
30468	625	80.3	80.9	81.6	8		
30478	632	69.9	70.4	75			

UCI Conte Center on Brain Programming in Adolescent Vulnerabilities (contecenter.uci.edu)
Dave Keator
Comprehensive and long lasting markup of a complex dataset



FreeSurfer Domain Object



David Keator et al., Neuroimage, 2013

Use case 2: FreeSurfer

Example of converting single files

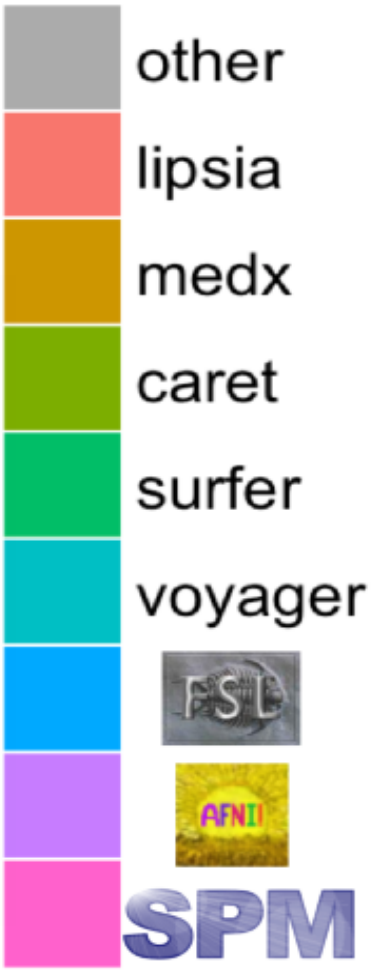
```
In [5]: g = prov.ProvBundle()
        e1 = create_entity(g, "bert", os.path.join(sdir, "mri/T1.mgz"))
        e2 = create_entity(g, "bert", os.path.join(sdir, "label/lh.BA6.label"))
        fsdir_collection = g.collection(niiri['foo'])
        g.hadMember(fsdir_collection, e1)
        print g.rdf().serialize(format='turtle')
```

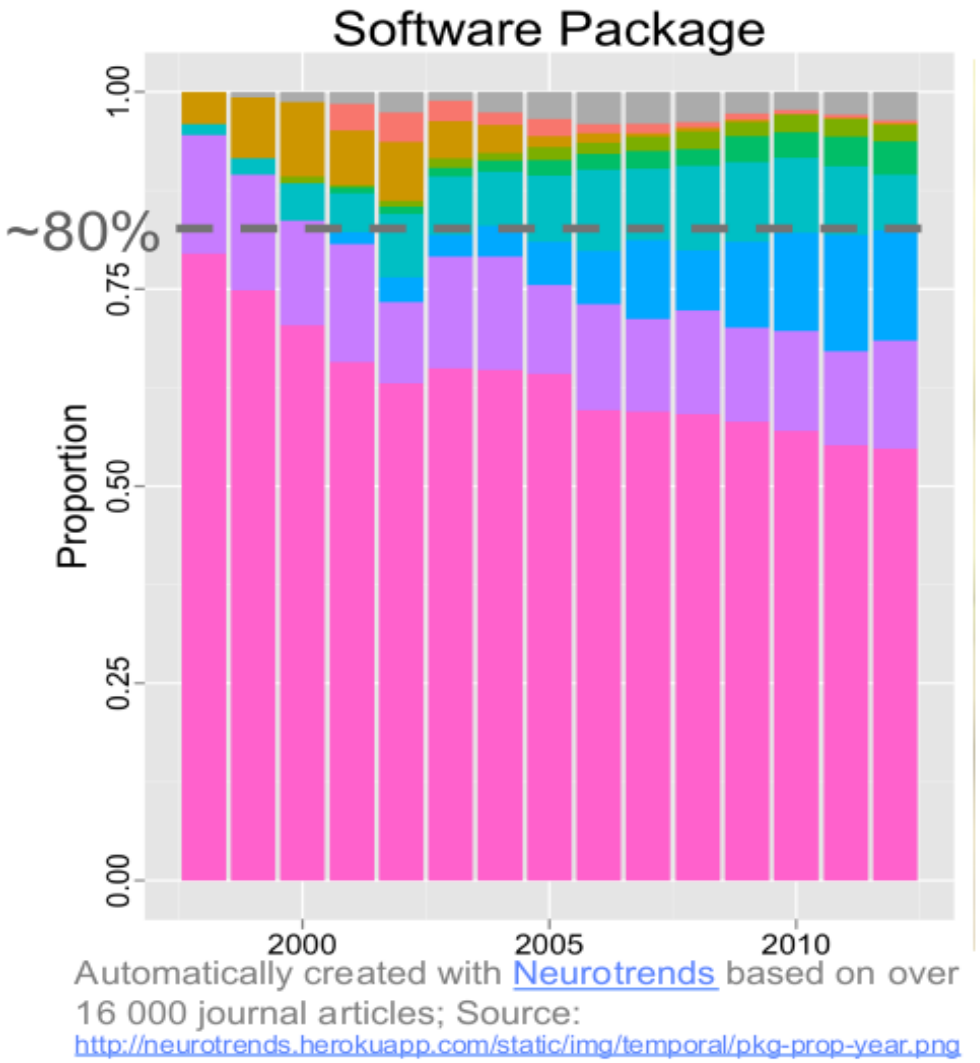
```
@prefix crypto: <http://www.w3.org/2000/10/swap/crypto#> .
@prefix fs: <http://freesurfer.net/fswiki/terms/0.1/> .
@prefix nidm: <http://nidm.nidash.org/terms/0.1/> .
@prefix nif: <http://neurolex.org/wiki/> .
@prefix niiri: <http://nidm.nidash.org/iri/> .
@prefix obo: <http://purl.obolibrary.org/obo/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
```

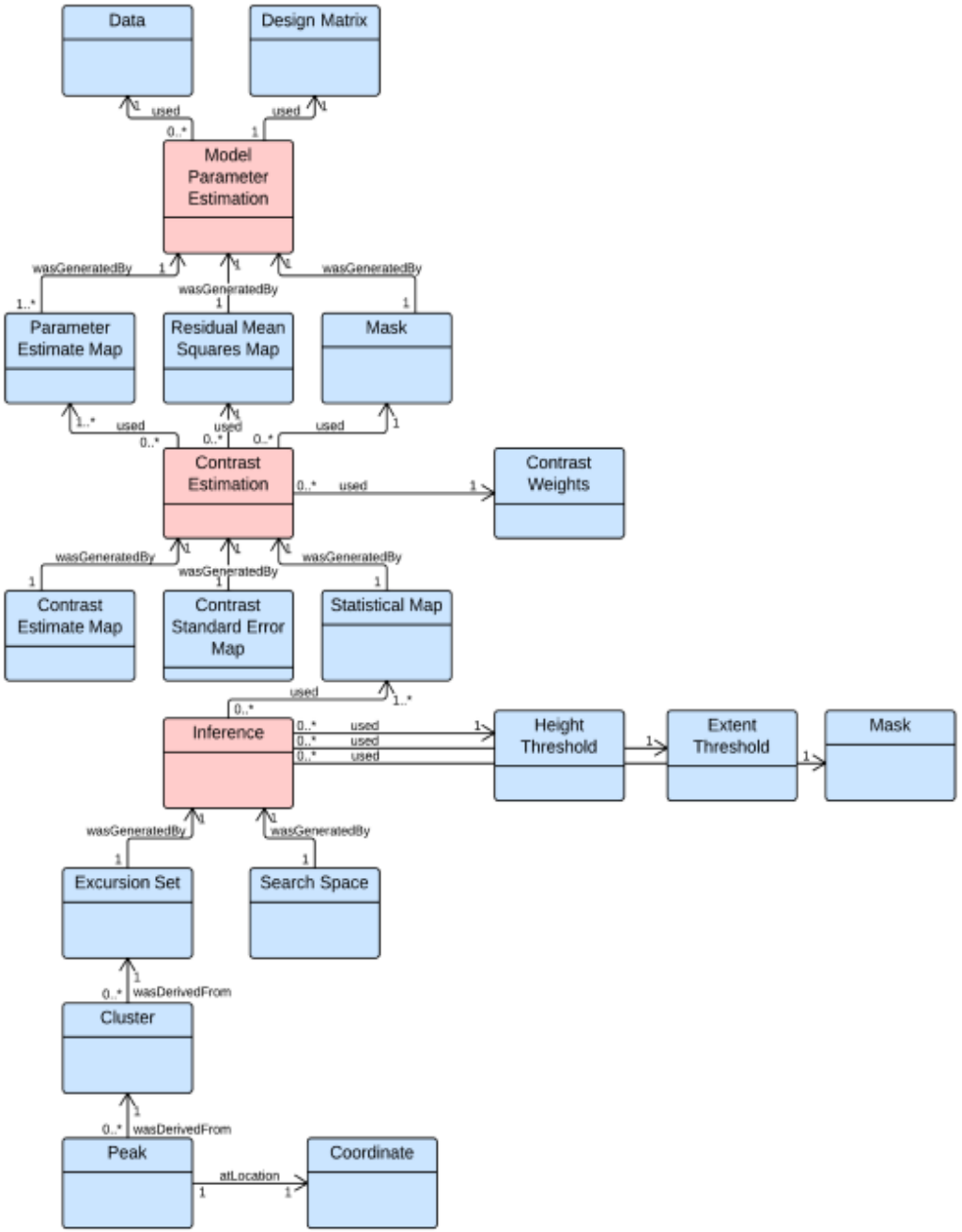
- Link to FMA
- Link to other tools
- Link to other data
- Github incf-NIDASH

Credit : Satra Ghosh

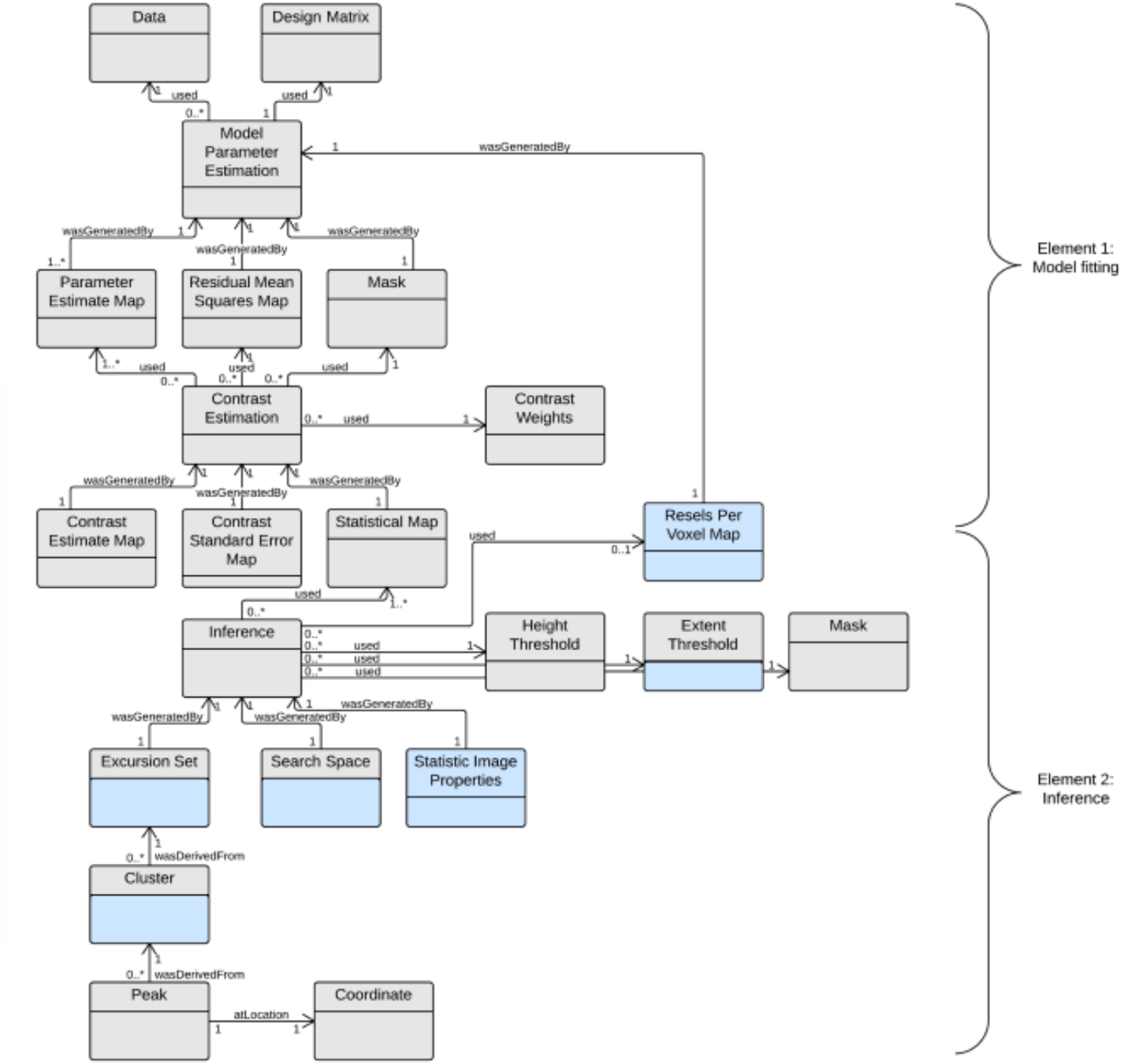
Use Case 3: fMRI statistical results

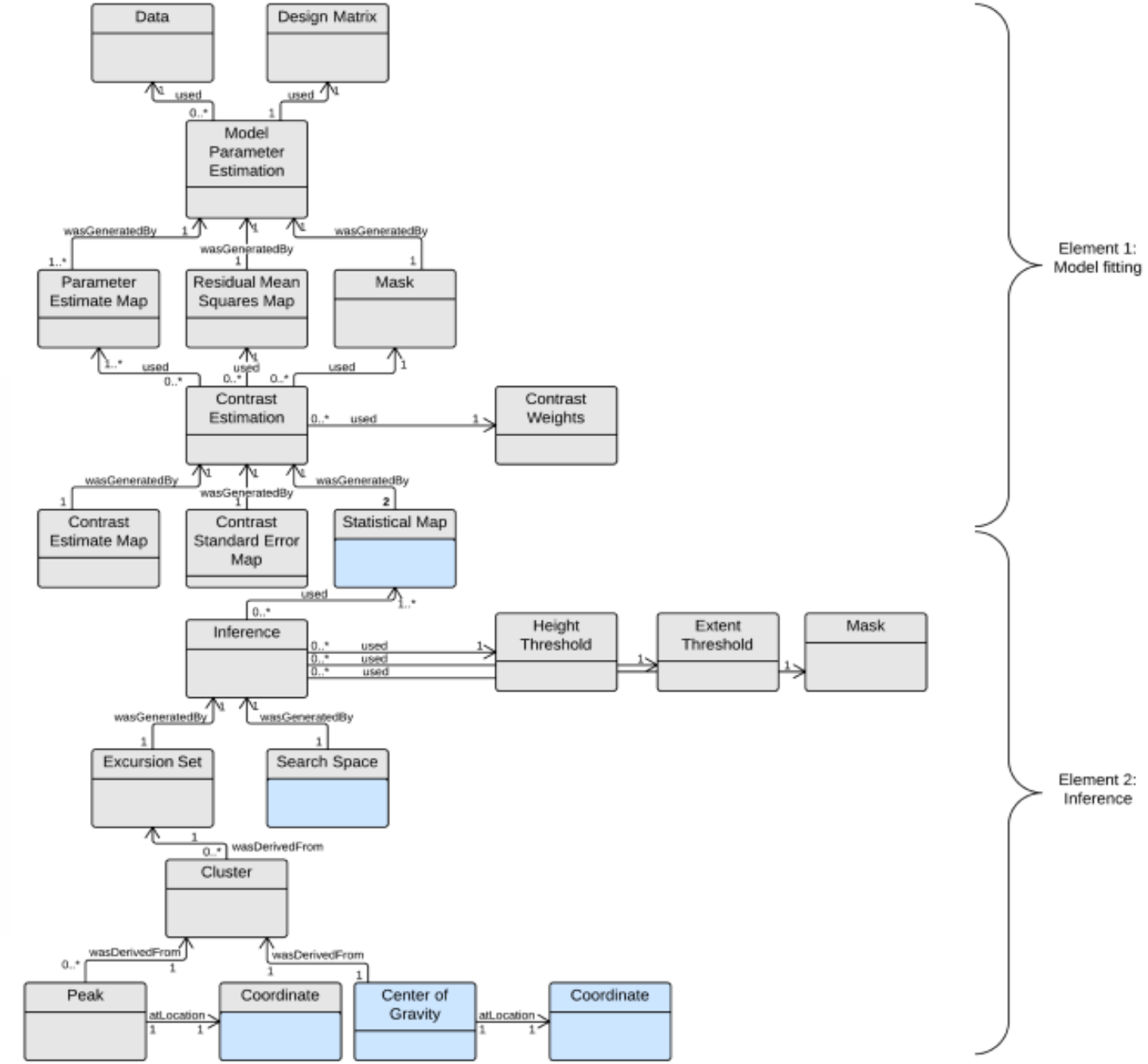





Data
Design Matrix
used
used
Model Parameter Estimation
wasGeneratedBy
wasGeneratedBy
wasGeneratedBy
Parameter Estimate Map
Residual Mean Squares Map
Mask
used
used
used
Contrast Estimation
Contrast Weights
used
wasGeneratedBy
wasGeneratedBy
wasGeneratedBy
Contrast Estimate Map
Contrast Standard Error Map
Statistical Map
used
Inference
used
used
used
Height Threshold
Extent Threshold
Mask
wasGeneratedBy
wasGeneratedBy
Excursion Set
Search Space
wasDerivedFrom
Cluster
wasDerivedFrom
Peak
atLocation
Coordinate
Element 1: Model fitting
Element 2: Inference


Design of the experiment, pre-processed data...
Statistical estimation
Effects, standard error and statistical maps...
Inference
Thresholded maps, list of peaks and clusters...


Data
Design Matrix
used
used
Model Parameter Estimation
wasGeneratedBy
Parameter Estimate Map
Residual Mean Squares Map
Mask
Contrast Estimation
Contrast Weights
Contrast Estimate Map
Contrast Standard Error Map
Statistical Map
Resels Per Voxel Map
Inference
Height Threshold
Extent Threshold
Mask
Excursion Set
Search Space
Statistic Image Properties
Cluster
wasDerivedFrom
Peak
atLocation
Coordinate
Element 1: Model fitting
Element 2: Inference

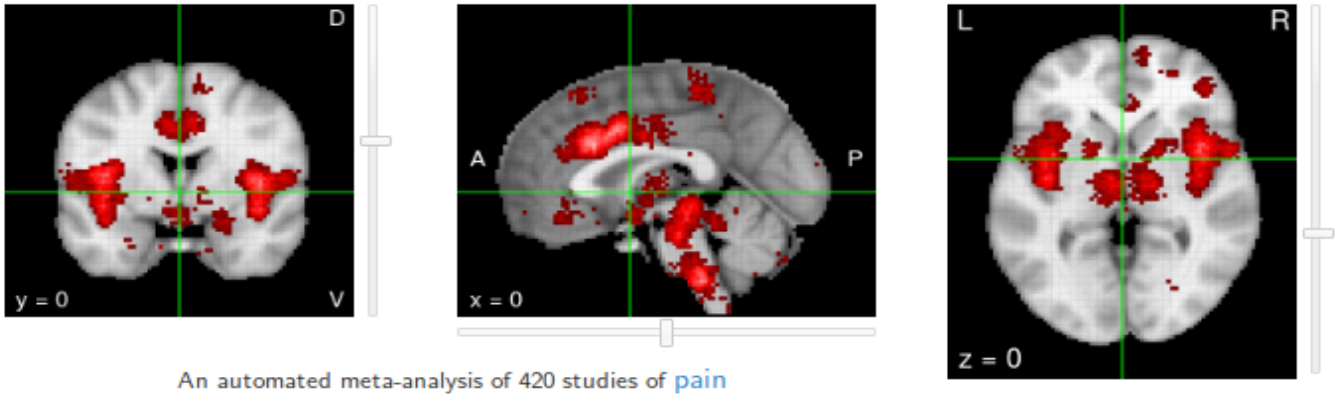

Data
Design Matrix
used
used
Model Parameter Estimation
wasGeneratedBy
wasGeneratedBy
wasGeneratedBy
Parameter Estimate Map
Residual Mean Squares Map
Mask
used
used
used
Contrast Estimation
Contrast Weights
used
wasGeneratedBy
wasGeneratedBy
wasGeneratedBy
Contrast Estimate Map
Contrast Standard Error Map
Statistical Map
used
Inference
Height Threshold
Extent Threshold
Mask
used
used
used
wasGeneratedBy
wasGeneratedBy
Excursion Set
Search Space
wasDerivedFrom
Cluster
wasDerivedFrom
wasDerivedFrom
Peak
atLocation
Coordinate
Center of Gravity
atLocation
Coordinate
Element 1: Model fitting
Element 2: Inference

Usual (x,y,z) meta analyses

neurosynth.org

Neurosynth is a platform for large-scale, automated synthesis of functional magnetic resonance imaging (fMRI) data.

It takes thousands of published articles reporting the results of fMRI studies, chews on them for a bit, and then spits out images that look like this:



An automated meta-analysis of 420 studies of pain

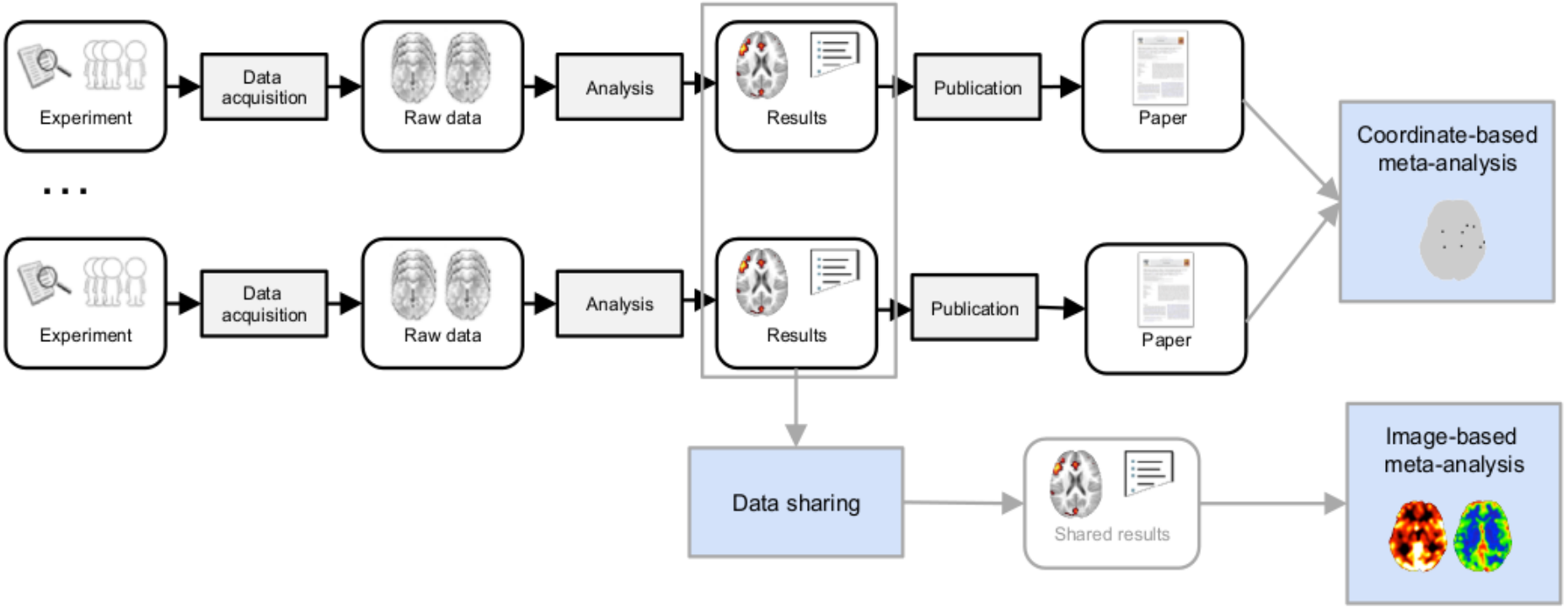
Database Status

413429 activations reported in 11406 studies

Interactive, downloadable meta-analyses of 3107 terms

Tal Yarkoni

Meta analyses - SPM+FSL(+Afni)



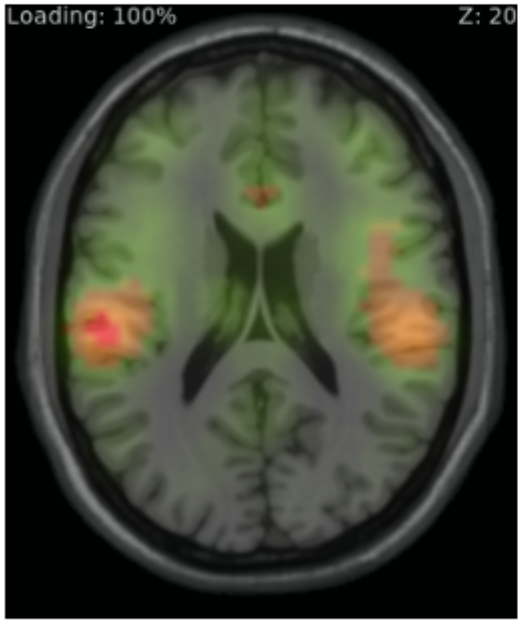
Smoothness, Error model, Contrast direction, ... marked up

brainspell

LOG IN BLOG ABOUT DOWNLOAD

pain

9179 articles indexed



Tags :
* **Cog Atlas**
* **CogPo**
* **Demogr.**
* **ROI**

368 articles corresponding to the search "pain"

Storing contrast maps and meta data

NeuroVault

A public repository of unthresholded brain activation maps

What is it?

A place where researchers can publicly store and share unthresholded statistical maps produced by MRI and PET studies.

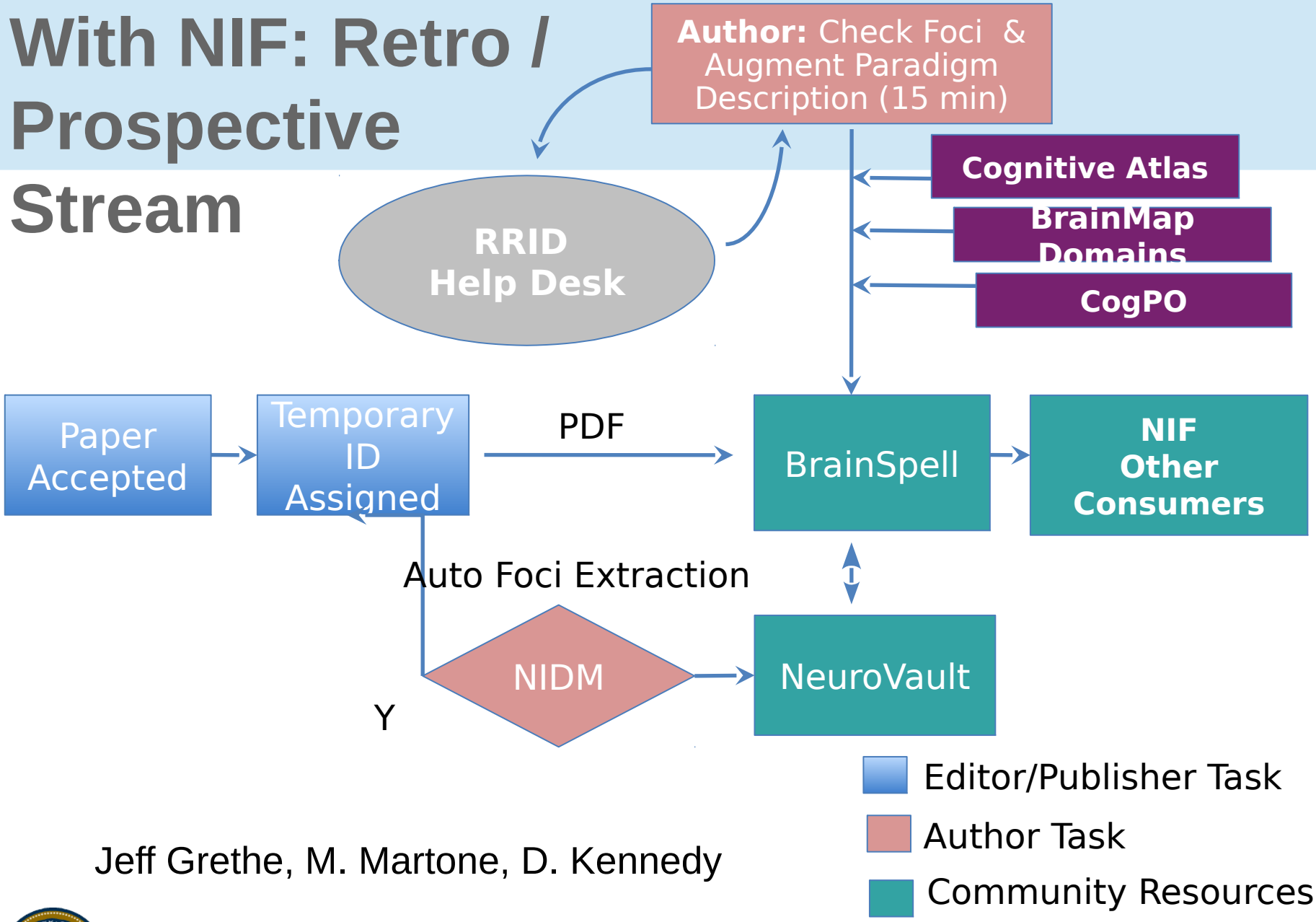
Why use it?

- Interactive visualization
- A permanent URL
- Publicly shareable
- Improves meta-analyses

Supported by

Chris Gorgolewski

With NIF: Retro / Prospective Stream



Jeff Grethe, M. Martone, D. Kennedy

Queries: where power will be demonstrated

- Standard SPARQL language
- Set of resources: local and/or distributed
- Create a common graph across resources

- Size of graphs ?
- Efficiency ?
- Tools: Virtuoso – 4store - Dydra - Talis:
  - Manage increasing large number of triples
- Training + Tools for neuroscientists

incf

Queries

In [12]:
```
qres = gmap.query(
    """SELECT DISTINCT ?subclasses
       WHERE {
              ?subclasses rdfs:subClassOf nidm:FSIQ .
              }""")
print qres.serialize(format='csv')
```

```
subclasses
http://fcon_1000.projects.nitrc.org/indi/abide/ABIDE_FIQ
http://www.birncommunity.org/collaborators/function-birn/FSIQ
```

Add the Mappings to the Joinined Datasets

In [15]:
```
mapped_graph = union_graph + gmap
```

Now use the mappings to filter the results!

In [17]:
```
qres = mapped_graph.query(
    """SELECT DISTINCT ?s ?subclasses ?o
       WHERE {
              ?subclasses rdfs:subClassOf nidm:FSIQ .
              ?s ?subclasses ?o
              }""")
print qres.serialize(format='csv')
```

```
s,subclasses,o
http://purl.org/nidash/nidm/entity_CMU_50645,http://fcon_1000.projects.nitrc.org/ind
i/abide/ABIDE_FIQ,124
http://purl.org/nidash/nidm/entity_CMU_50643,http://fcon_1000.projects.nitrc.org/ind
i/abide/ABIDE_FIQ,123
http://purl.org/nidash/nidm/entity_CMU_50647,http://fcon_1000.projects.nitrc.org/ind
i/abide/ABIDE_FIQ,104
http://purl.org/nidash/nidm/entity_CMU_50646,http://fcon_1000.projects.nitrc.org/ind
i/abide/ABIDE_FIQ,108
```

Queries -

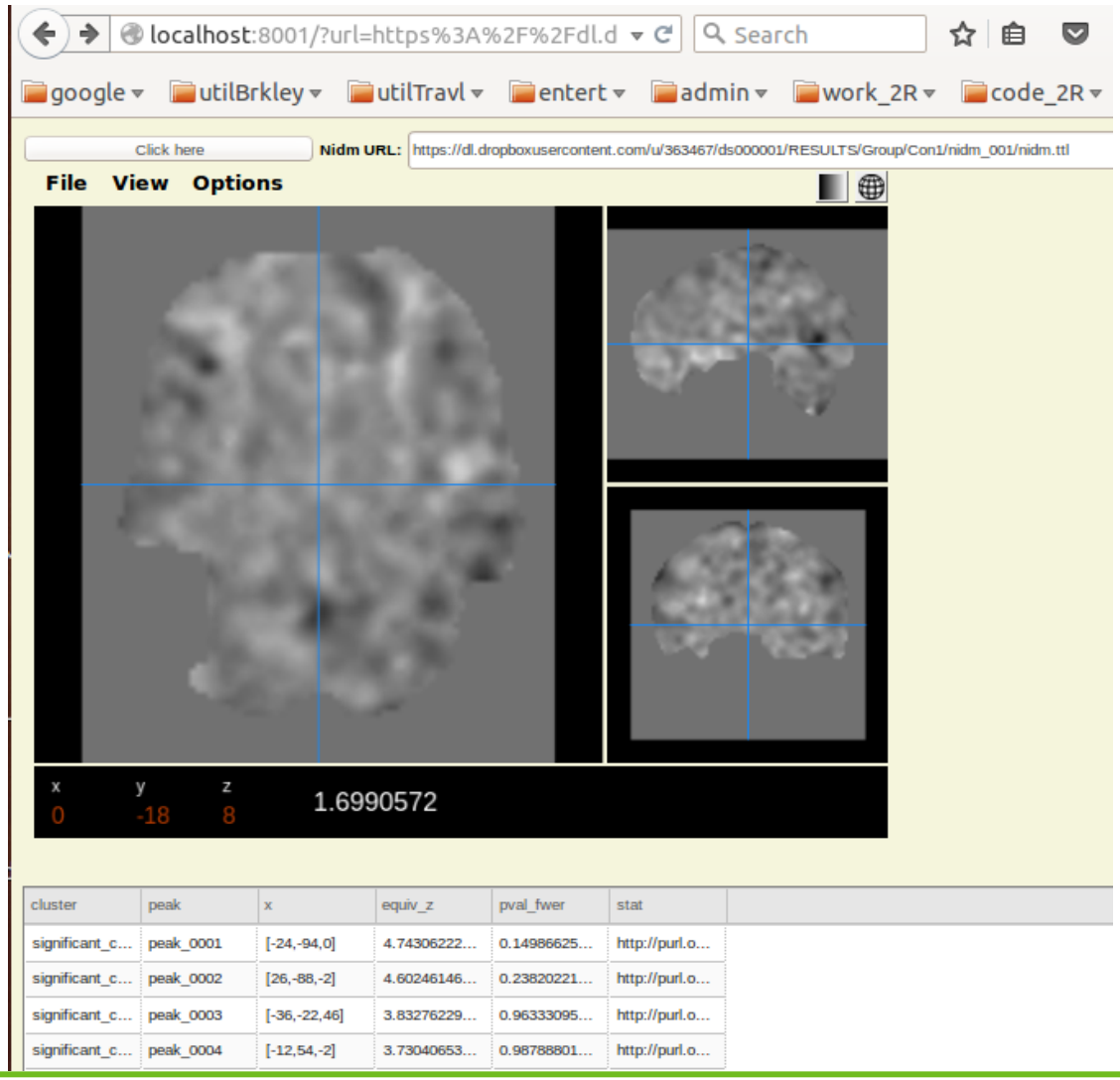
```
SELECT DISTINCT *
WHERE
{ ?peak a nidm:Peak .
  ?cluster a nidm:Cluster .
  ?peak prov:wasDerivedFrom ?cluster .
  ?peak prov:atLocation ?coordinate .
  ?coordinate nidm:coordinate1 ?x .
  ?coordinate nidm:coordinate2 ?y .
  ?coordinate nidm:coordinate3 ?z .
  OPTIONAL { ?peak prov:value ?value }.
  ?peak nidm:equivalentZStatistic ?zstat .
  OPTIONAL { ?peak nidm:pValueFWER ?pvalcor }.
  ?peak nidm:pValueUncorrected ?pvalunc .
  ?cluster
prov:wasDerivedFrom/prov:wasGeneratedBy/prov:us
ed ?statmap .
  ?statmap a nidm:StatisticMap .
  ?statmap nidm:statisticType ?stat .
}
ORDER BY ?cluster ?peak
```

¶['http://iri.nidash.org/cluster_0001', '4.126074e-10', 'http://iri.nidash.org/peak_0001', 'http://iri.nidash.org/coordinate_0001', '-48.1', '-9.24', '-73.7', 'http://www.incf.org/ns/nidash/nidm#ZStatistic', 'http://iri.nidash.org/z_statistic_map_id', 'None', '6.14', 'None']

¶['http://iri.nidash.org/cluster_0001', '7.705712e-10', 'http://iri.nidash.org/peak_0002', 'http://iri.nidash.org/coordinate_0002', '-38.1', '-18.0', '-53.4', 'http://www.incf.org/ns/nidash/nidm#ZStatistic', 'http://iri.nidash.org/z_statistic_map_id', 'None', '6.04', 'None']

¶['http://iri.nidash.org/cluster_0001', '4.462172e-09', 'http://iri.nidash.org/peak_0003', 'http://iri.nidash.org/coordinate_0003', '-29.6', '-16.9', '-73.8', 'http://www.incf.org/ns/nidash/nidm#ZStatistic', 'http://iri.nidash.org/z_statistic_map_id', 'None', '5.75', 'None']

NIDM-results viewer



NIDM-Results^2

- FS converter
- FSL converter
- Native SPM exporter
- AFNI engaged in the project

- OpenFMRI data nidm results for both SPM and FSL
- Neurovault ingests nidm results
- BIDS to NIDM soon

- Nidm used in Conte, NCANDA, NKI soon & other projects

Four grants have been submitted to support some aspects of NIDM

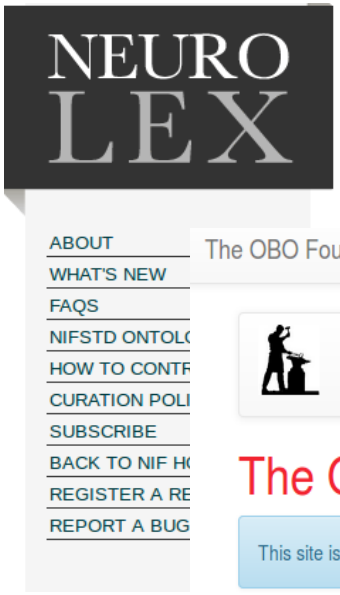
incf

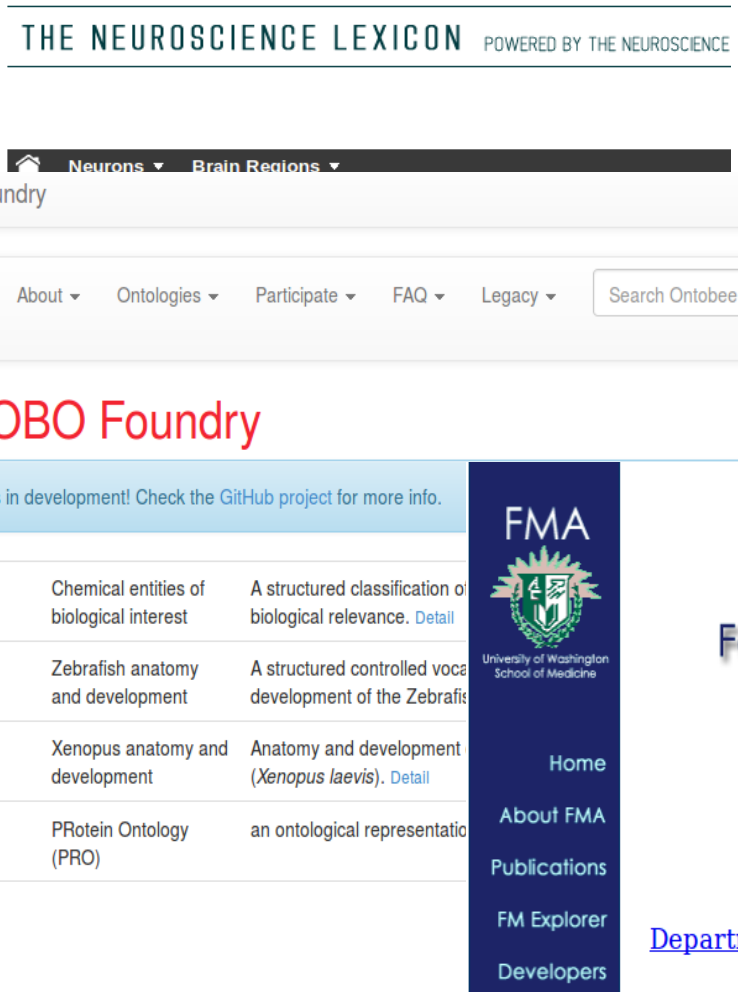
How do we understand each others ?

- We need a common goal
- We need a common grammar (NIDM)
- We need a common vocabulary
- We need a way to re-use “terms” and augment the vocabulary

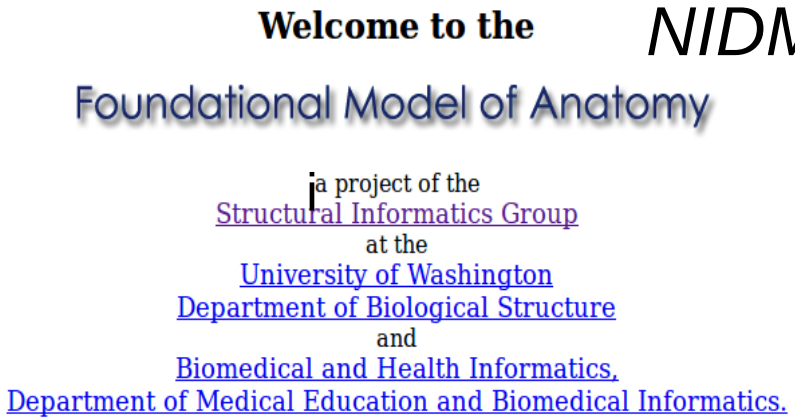
→ **NIDM: A workflow to curate terms openly**

Finding terms: where should I look





Cognitive Atlas



DCT
OBO
RDFS
HCSI
NCIT
STATO
NIF
NIDM

Why this is –may be- working ?

- Spirit based on open source development movement has proven to be efficient – get inspired by Open Science Foundation type of initiative
- Based on a strong neuroimaging and python open source ecosystem:
    - Nipy, Nipype, NiTime, Dipy, Nilearn, MNE, etc
- Because we are trying to solve a problem in a domain we know well (500 years of experience – and yet some fresh brains)
- Funding: opportunity with NIDASH / INCF

Because of the individuals in the group

incf

What are the difficulties?

- Coordination: short vs long terms goals
- Software development requires sustained resources
- Small developer community
- Software development: PhDs and grants not common – publications focus - but see Giga science
- Most researchers have domain knowledge, but not the technical skills
- Steep learning curve to understand all technologies involved: need for training

incf

Current/Future work

- NIDM experiment:
  - Many more datasets NKI, NIDM-BIDS
  - Link with NDAR
- LORIS, XNAT, COINS, HID, etc: engage to read and export NIDM-experimental – Common API
- NIDM-Workflow :
  - Integrate Nipype, C-Pack, C-BRAIN, AA, etc, with a common provenance model
  - A common language for processing
- NIDM-Results:
  - Other software, methods (resting state networks....)

incf

Acknowledgements

TO MATHEW, ROSA, EVA-CHRISTINE, LINDA
And all at the INCF Secretariat