Information Integration and Knowledge Engineering at BIRN

BIRN
Biomedical Informatics
Research Network

Jessica Turner

Mind Research Network, NM

OBI Workshop — March 22, 2010

Biomedical Informatics Research Network

Goal: Use a bottom-up approach to advance biomedical research through data sharing and online collaboration.

Services: data-sharing infrastructure, software tools, strategies and advice to individual biomedical research users and user groups.

BIRN Biomedical Informatics Research Network

BIRN collaborating projects

- Functional Imaging BIRN
  - HID
  - XNAT
  - Northwestern Schizophrenia project
- Mouse BIRN
- Cardiovascular Research Grid
- Nonhuman Primate Research Consortium
- HEAD IT (EEG data)
- Cognitive Paradigm Ontology

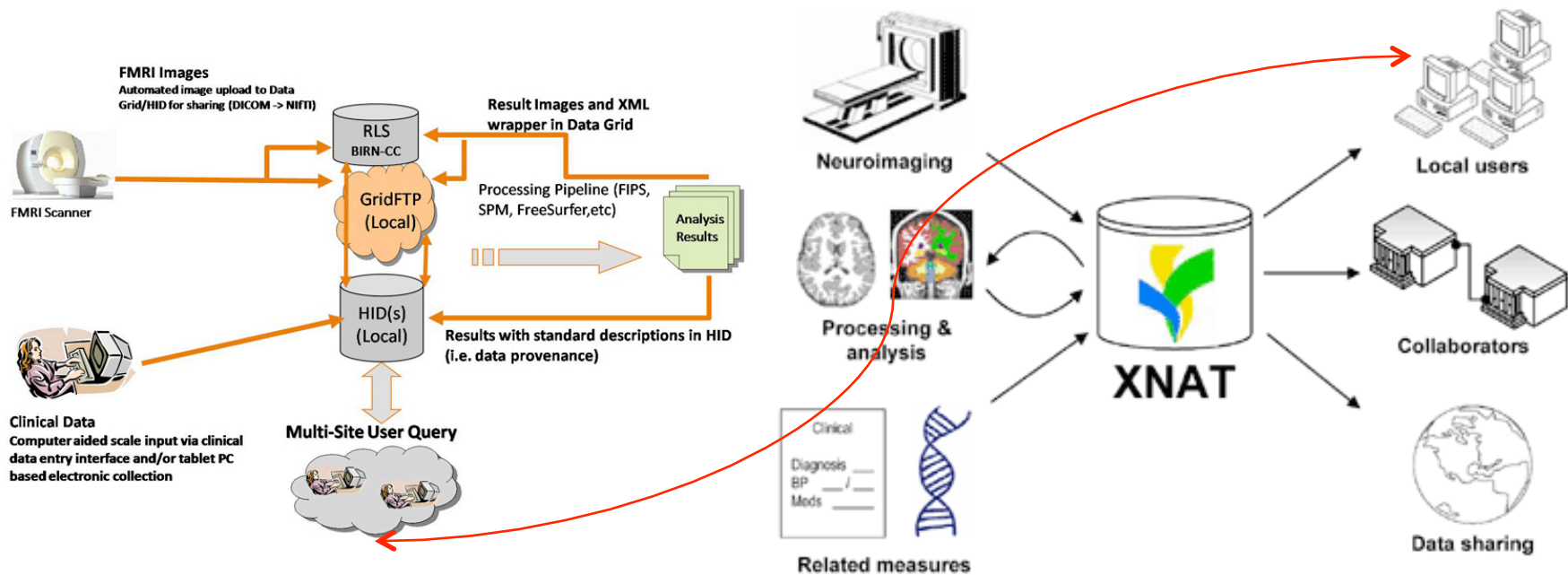BIRN Biomedical Informatics
Research Network

KEWG Approach

- Lightweight methods focused on scalability and ease-of-use
- Orthogonal / Complementary to efforts within ontology curation community
  - effort to not reproduce work done within previous BIRN
- Centered on two technical components
  - Knowledge Engineering from Experimental Design ('KEfED')
  - Ontology of Experimental Variables and Values ('OoEVV')

BIRN Biomedical Informatics Research Network

KEWG Approach

- Lightweight methods focused on scalability and ease-of-use
- **Orthogonal / Complementary to efforts within ontology curation community**
  - effort to not reproduce work done within previous BIRN
- Centered on two technical components
  - Knowledge Engineering from Experimental Design ('KEfED')
  - Ontology of Experimental Variables and Values ('OoEVV')

BIRN Biomedical Informatics Research Network

KEWG Approach

- Lightweight methods focused on scalability and ease-of-use
- Orthogonal / Complementary to efforts within ontology curation community
  - effort to not reproduce work done within previous BIRN
- Centered on two technical components
  - Knowledge Engineering from Experimental Design ('KEfED')
  - Ontology of Experimental Variables and Values ('OoEVV')

BIRN Biomedical Informatics Research Network

Information Integration

Information Integration: Retrieving from multiple sources

Information Integration

Example of information integration query

- *Find all the subjects who have both a structural and functional MRI scan*

- Q(subjectid) :- g_ExperimentAcquisition(_, subjectid, scantype1, _, _, _, _, _)
  - ^ g_ExperimentAcquisition(_, subjectid, scantype2, _, _, _, _, _)
  - ^ (scantype1 LIKE "structural MRI")
  - ^ (scantype2 LIKE "functional MRI")

Example of information integration query

- *Find all the subjects who have both a structural and functional MRI scan*

```
g_ExperimentAcquisition(acqID, projectID, "NA", acqsessiontype, acqscantype,
                         subjectID, "NA", "NA", "NA", acqscanner, acqmarker)
      :- XnatMRSessionResource_xnat__mrSessionData(acqID, subjectID, XNAT_DATE,
                                INVEST_SEARCH, INVEST, OPERATOR,
                                acqsessiontype, acqscanner, acqmarker,
                                STABILIZATION, OPERATOR_CSV, SCANNER_CSV,
                                MARKER_CSV, STABILIZATION_CSV, AGE, DERIVED_AGE,
                                INSERT_DATE, INSERT_USER, LAST_MODIFIED,
                                MR_SCAN_COUNT_AGG, PROJECTS, projectID,
                                SHARED_PROJECT, LABEL, DCMACCESSIONNUMBER, DCMPATIENTID,
                                DCMPATIENTNUMBER, DCMUID, acqscantype)


g_ExperimentAcquisition(acqID, projectID, episodeID, "NA", acqscantype,
                         subjectID, acqtimeduration, acqprotocol, acqstudy, "NA", "NA")
      :- HidResource_NC_EXPSEGMENT(acqID,episodeID,projectID,subjectID,
                                uniqueid,tableid,owner,modtime,moduser,time_stamp,
                                description,protocolversion,protocolid,studyid,acqscantype,
                                istimeinterval,acqtimeduration,isbad)
      ^ HidResource_NC_PROTOCOL(PROTOCOLVERSION2,protocolid,UNIQUEID2,TABLEID2,OWNER2,
                                MODTIME2,MODUSER2,acqprotocol,DESCRIPTION2)
      ^ HidResource_NC_EXPSTUDY(studyid,COMPONENTID3,EXPERIMENTID3,SUBJECTID3,UNIQUEID3,
                                TABLEID3,OWNER3,MODTIME3,MODUSER3,TIME_STAMP3,DESCRIPTION3,
                                acqstudy,ISTIMEINTERVAL3,TIMEINTERVAL3)
```

Information Integration

- Develop common domain model

- Use ontology terms to construct domain model

- Map each database structure to model

- Integrate via common domain model

BIRN Biomedical Informatics Research Network

Example: Phenotype Qualities

Variable		Possible Values
• Age	PATO_0000011	*(years)*
• Gender	PATO_0001894	"male", "female"
• Handedness	birnlex_2178	Edinburgh Handedness Inventory score
• Smoking Status	birnlex_2102 (?)	Number of pack-years, "yes"/ "no"
• Alzheimers		Clinical Dementia Rating score
• Mental State		Mini Mental State Examination score
• Clinical severity		Scale for Assessment of –/+ symptoms
• Diagnosis		*(list of diagnoses)*
• Scan Type		*(list of types)*

Vision of the new BIRN

- OOEVV/KEfED for simplistic, rapid prototyping of experimental methods

Experimental Type →

	e.g. lesion experiments	e.g. tract-tracing experiments	e.g. activation experiments
high-level interpretations ('punchline')		'brain region A projects to brain region B'	
primary experimental observations		'tracer A was injected into region B and labeling of type C was observed in regions D, E & F	
complete details of experimental methods & results		number of rats, type of injection, handling protocol, methods of data analysis, etc.	
nuanced representation of reliability		quality of histology, reputation of authors, etc.	

Depth of Reoresentation ↓

KEfED / OOEVV

OBI

Ontology of Experimental Variables and Values ('OoEVV')

- Lookup table for simple semantics of experimental variables
  - We want to compare variables that measure the same thing.
- Based on Measurement Scales (nominal, ordinal, interval, ratio, *etc.*[1])
  - *May need to extend Stevens' scales to include more complex elements like hierarchical taxonomie*
- Want to publish this to standardized ontology search engines (NCBO, OBO Foundry *etc.*)
  - Expressed in RDF / OWL.
  - Serves as a focal point of interaction with other ontology curation activities

[1] Stevens, S. S. (1946). "On the theory of scales of measurement." Science **103**(2684): 677-680.

OoEVV Basic Design
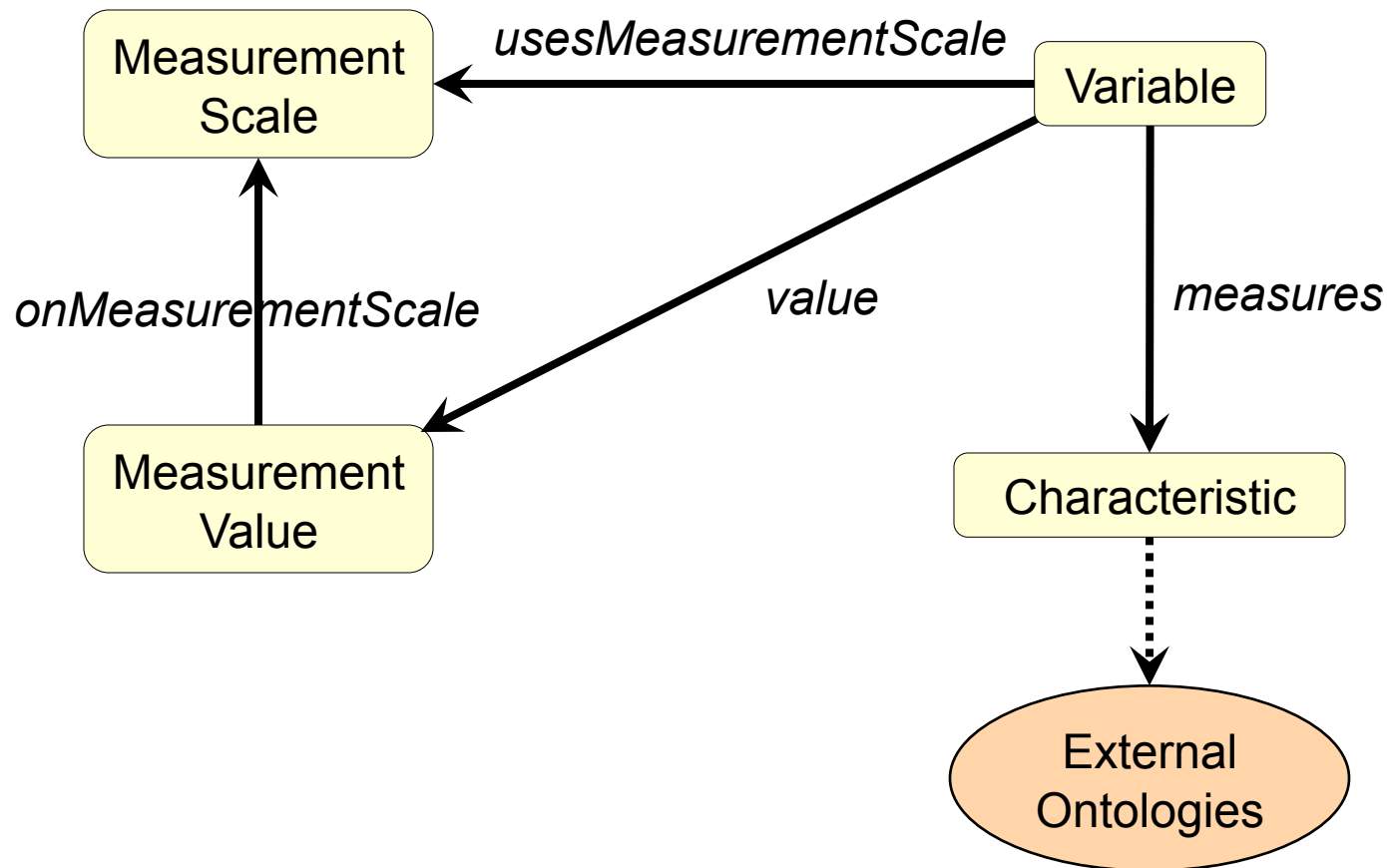
Example: Simple ('nominal') Handedness

Characteristic	Variable	Measurement Scale	Measurement Value
		Nominal Measurement Scale	Nominal Measurement Value

Handedness	Handedness Variable		Nominal Handedness Value
BirnLexOntology: birnlex_2178	Nominal Handedness Variable	Nominal Handedness Scale	left handed right handed ambidextrous

Example: Edinburgh Handedness Inventory

Characteristic

Variable

Measurement Scale

Measurement Value

Ordinal Measurement Scale

Ordinal Measurement Value

Handedness

Handedness Variable

Edinburgh Handedness Value

BirnLexOntogy: birnlex_2178

Edinburgh Handedness Variable

Edinburgh Handedness Inventory

float values [-100.0, +100.0]

Examples: Scales

OOEVV so far
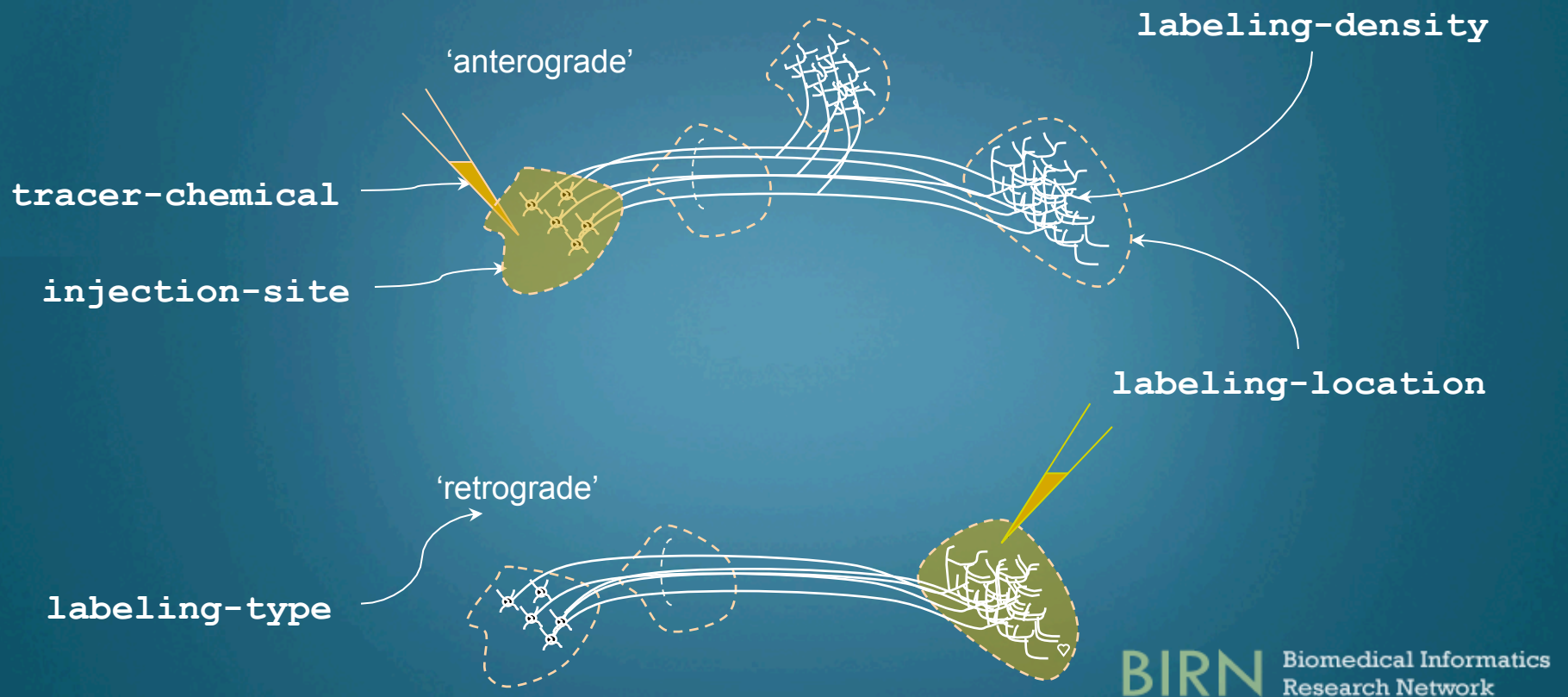
Example: Neural Connectivity - Experimental Observations

Tract Tracing Experiments

Neuroanatomical experiments to study neural connectivity.

labeling-density

'anterograde'

tracer-chemical

injection-site

labeling-location

'retrograde'

labeling-type

Example: Neural Connectivity - Interpretations

Tract Tracing Experiments > Neuroanatomical Elements
Interpretative entities that correspond to facts that may be aggregated into a model
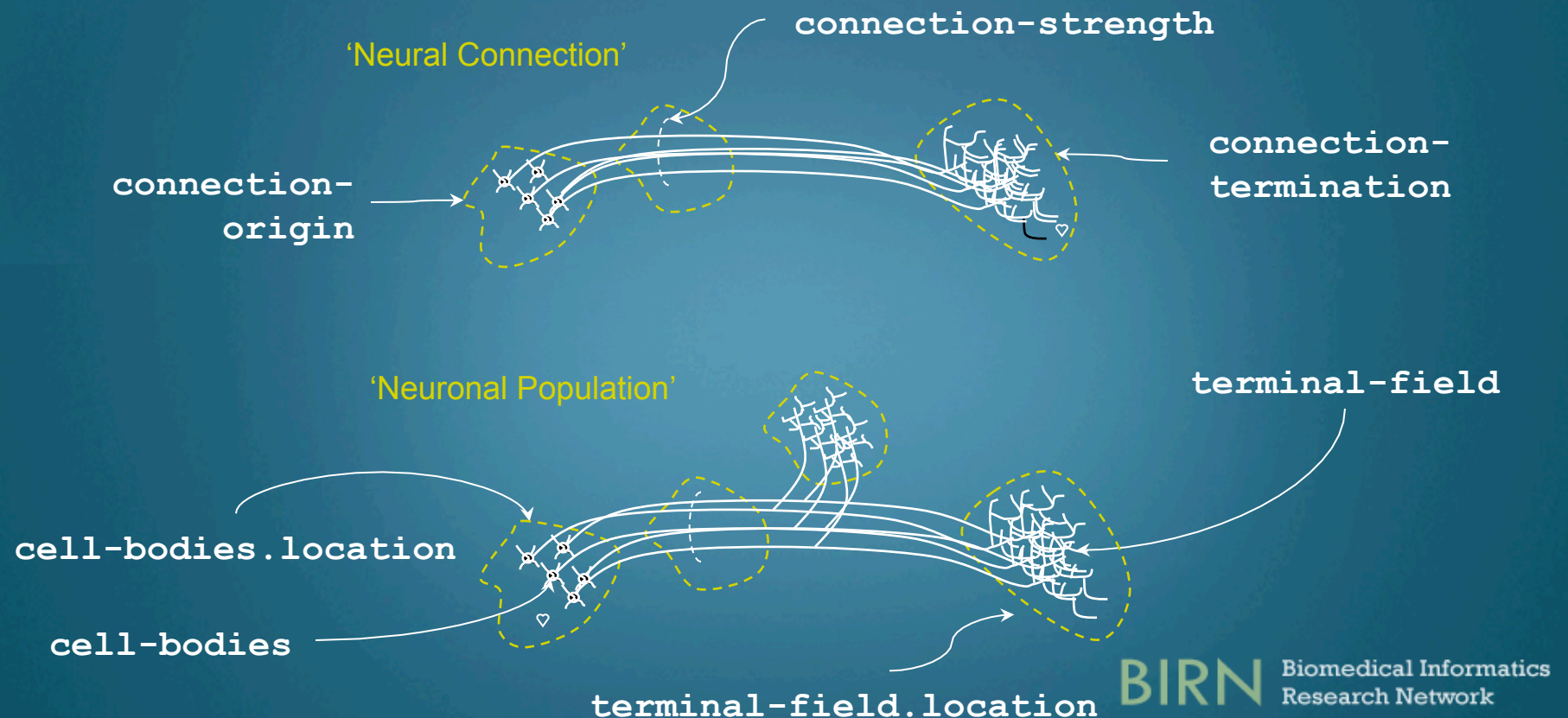
Knowledge Engineering from Experimental Design ('KEfED')

- There is an *implicit* reasoning model employed by scientists
  - Independent Variables *[experimental constraints]*
    - *Control Variables*
  - Dependent Variables *[measurements]*
  - Derived Variables *[calculations]*

- Interpretative statements are generated from data, relations and correlations between these variables
- How can we capture these relations within experiments?

BIRN Biomedical Informatics Research Network

KEfED Elements

Logical Element	Icon	OBI
Activity		Process
Experimental Object		Material Entity
Independent Variable		Independent variable specification
Dependent Variable		Dependent variable specification
Branch		
Fork		

BIRN Biomedical Informatics Research Network

Dependencies between variables are inherent in the experimental protocol
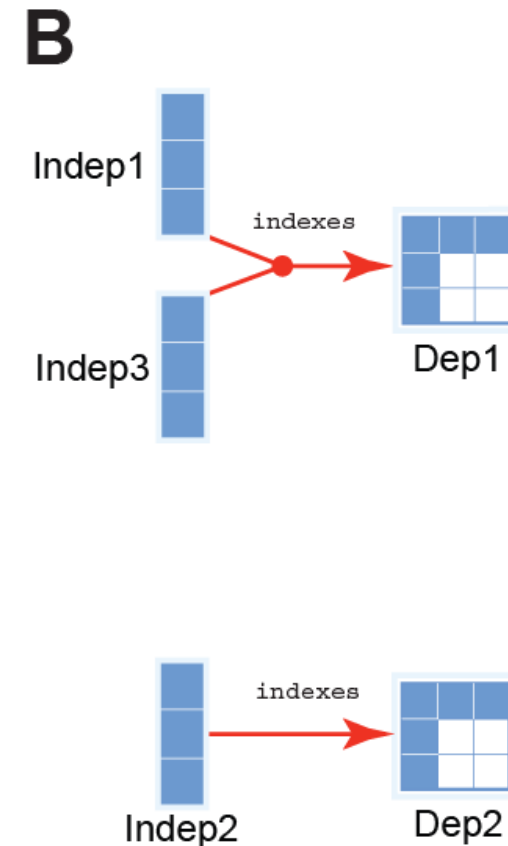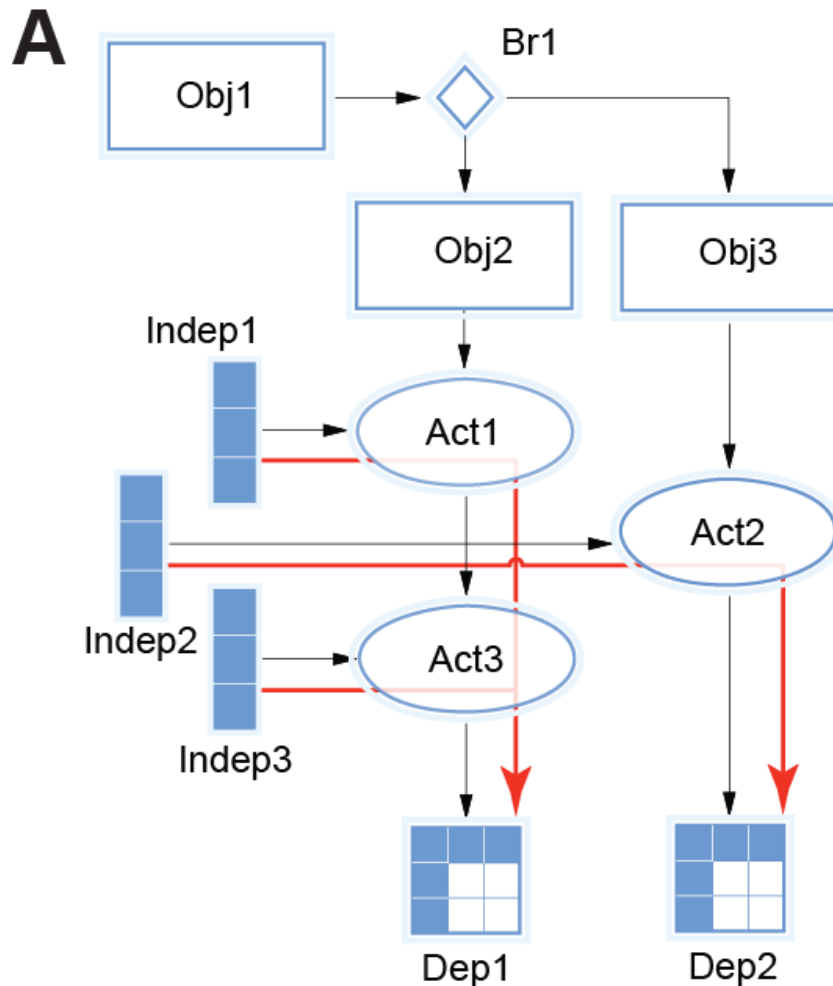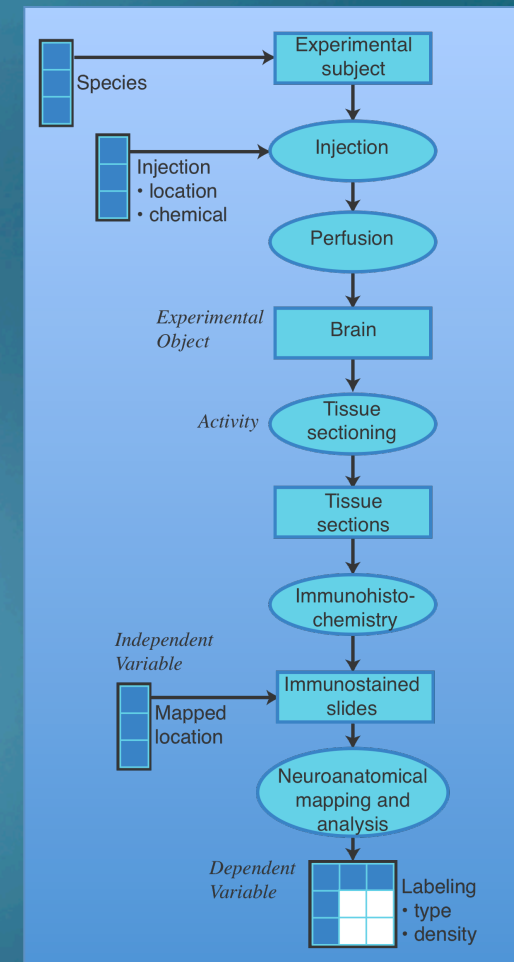
Experiment Modeling

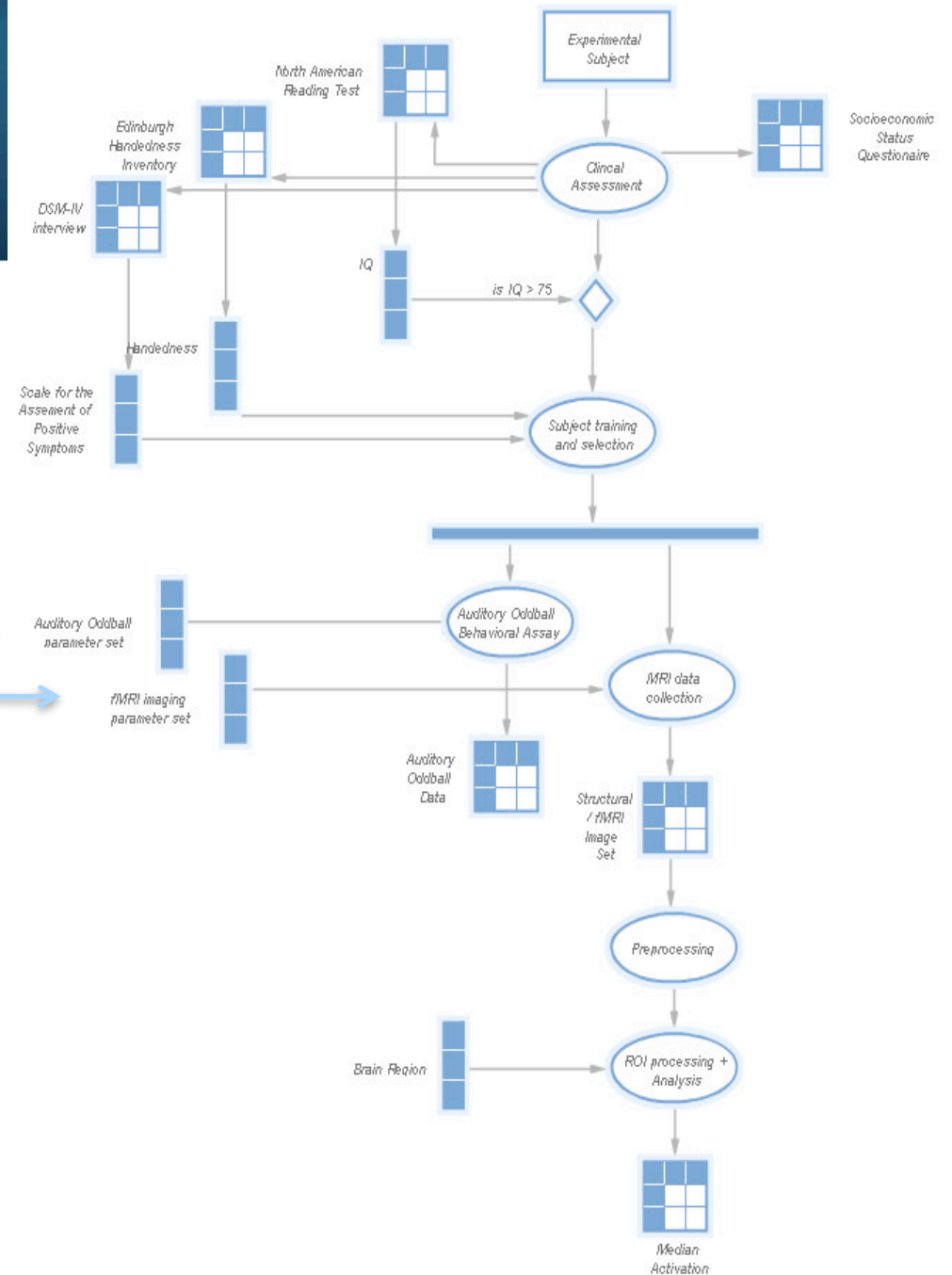- ## Describe Experimental Design as Process Model

  OBI, CogPO, NEMO, ?

- ## Data Modeled as Independent and Dependent Variables

  OBI, IAO, PATO, NeuroLex, ?

KEfED modeling of single subject fMRI study

CogPO
comes in here
RadLex here
Etc.

Other technology based on KEfED: 'Crux'

A semantically-enabled data repository designed to support meta-analysis across individual studies within a disease foundation

[http://neurosys.msu.montana.edu/Yogo/crux.html]

- Funded by the Kinetics and Michael J Fox Foundations
- Goal is to have an adminstrator design a KEfED model that can then be used as the basis for a data repository
- Twin challenges:
  - how can we build a data repository quickly for researchers?
  - how do we curate ontology terms accurately?
- Alan Ruttenberg is responsible for ontology development within project

BIRN Biomedical Informatics Research Network

KEfED models of Tract-tracing experiments in PowerLoom
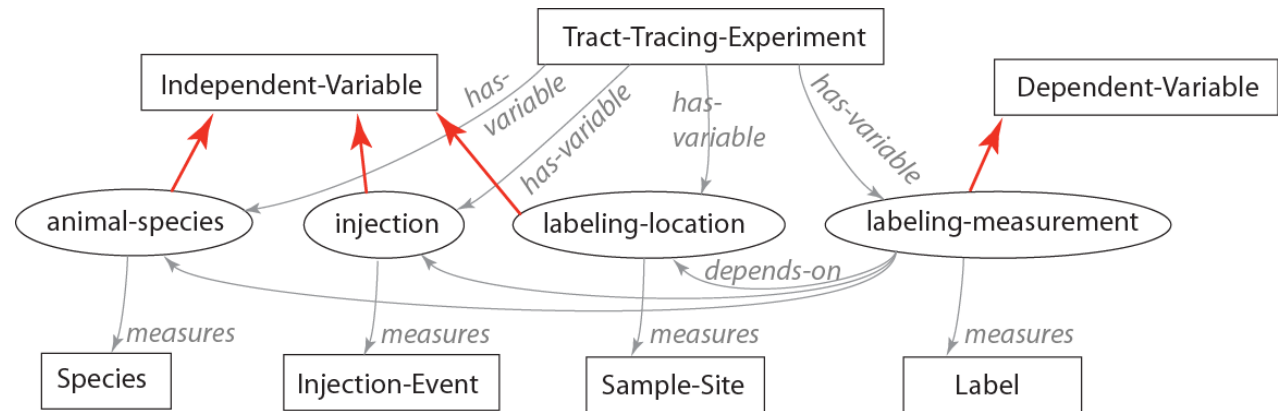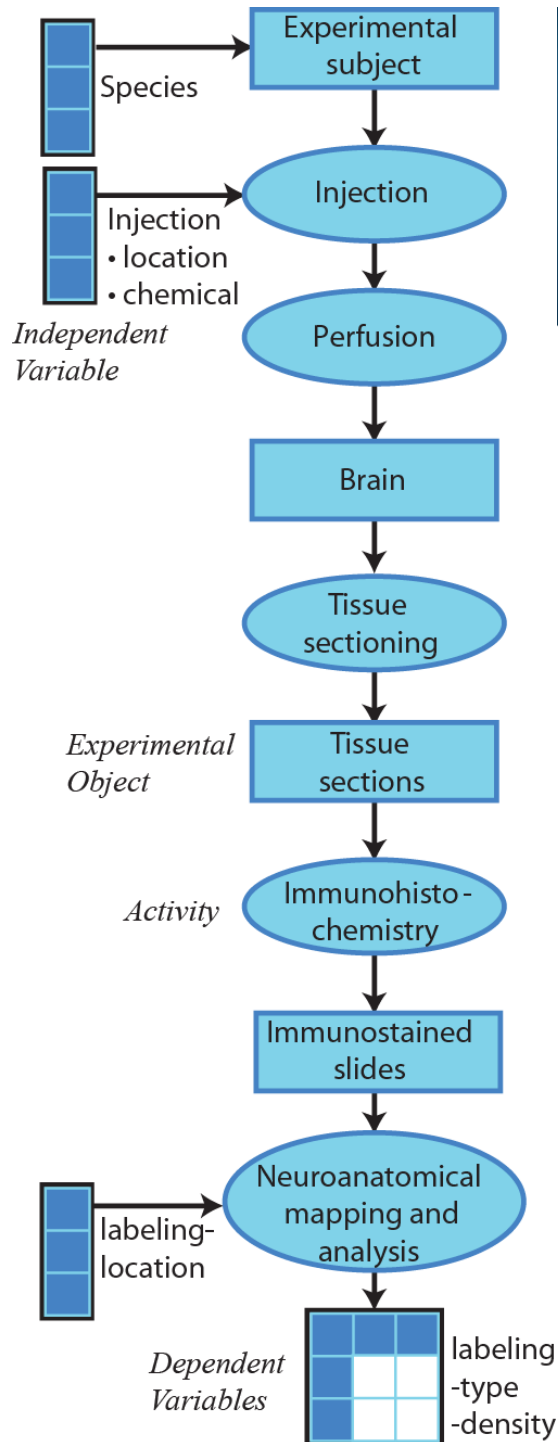
Assertion: injection into SUBv generates labeling in *MM*
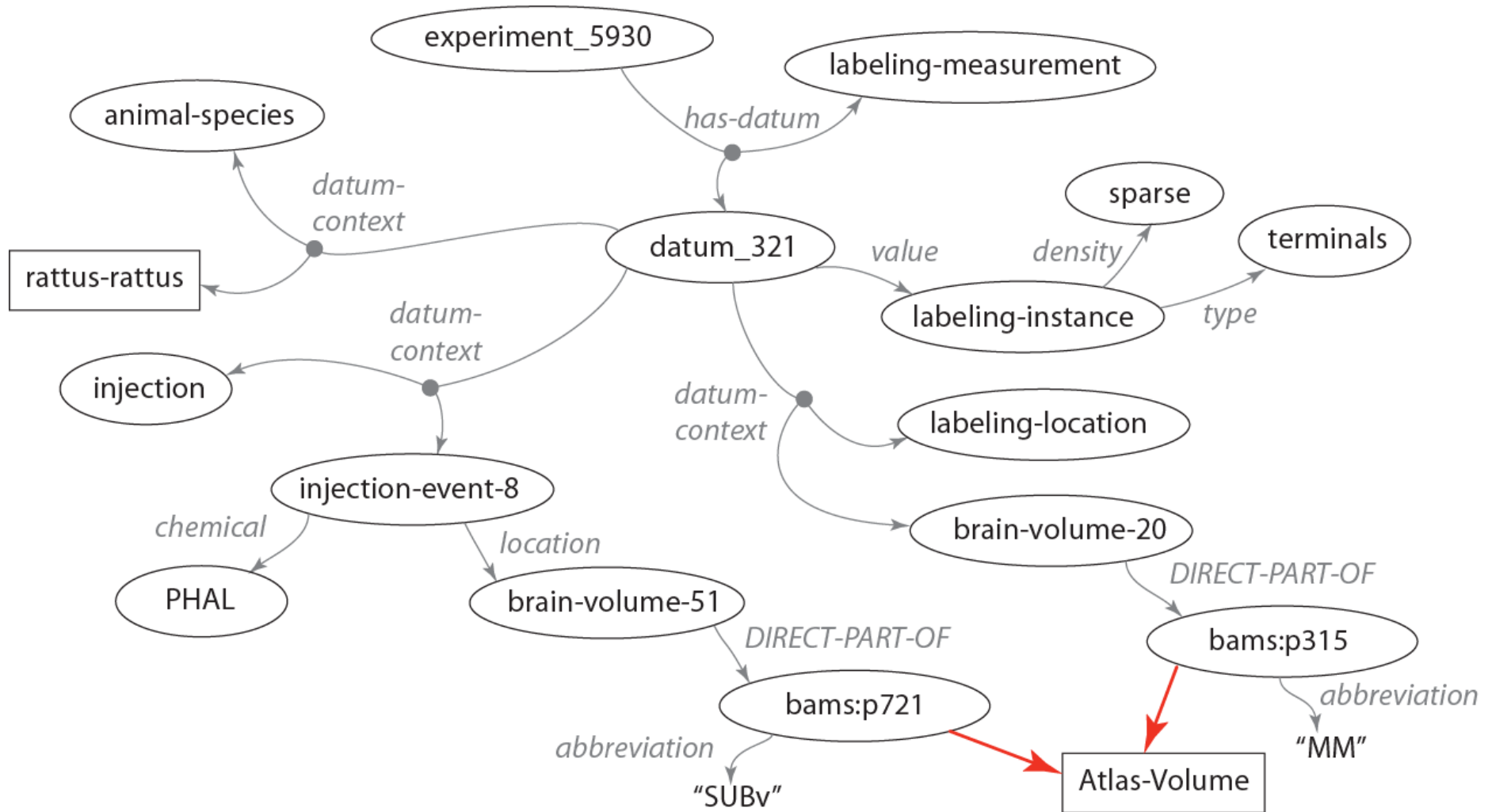
How to have this interoperate with the broader biomedical semantic web community?

Split the representation into OWL / PowerLoom components

- Ontology of Experimental Varriable and Values (OoEVV)
  - *e.g. handedness (can be measured by 'EHI' scale)*
  - Standard OWL ontology to be published in open community resources (OBO Foundry, NCBO, *etc.*)

OWL
PL
- KEfED : Models of Experimental Design + Data
  - Need more expressive reasoning
- Domain-specific reasoning models
  - *e.g. 'neural connectivity' models*

Goals

- Develop straightforward process for curation of variables into OoEVV
  - contribute these elements to OBI
- Make OoEVV entries for BIRN collaborators available to information integration work

BIRN Biomedical Informatics Research Network

Things We Need

- interaction between KEfED / OoEVV and OBI
  - A representation of measurement scales and associated values
  - Units and Measures representation
  - Links to specific terms for protocol steps, material entities and information entities
  - How to easily leverage OBI structure into our systems and software?

Teams + Collaborators

- BIRN KEWG Direct Members
  - Gully Burns
  - Jessica Turner
  - Tom Russ
  - Ed Hovy
  - *(Jose Luis Ambite)*
  - *(Naveen Ashish)*
- Biomedical Knowledge Engineering Group @ ISI
  - Cartic Ramakrishnan
  - Tommy Ingulfsen
  - Abhishek Patnia

Technology Collaborators

- Ontologies / KR&R
  - Alan Ruttenberg (ScienceCommons)
- NLP
  - Larry Hunter (Colorado)
  - Ellen Riloff (Utah)
  - Lynette Hirschmann (Mitre)
- Crux
  - Gwen Jacobs (Montana)
  - Michael Rogan (Kinetics Foundation)

BIRN Biomedical Informatics Research Network