Using knowledge engineering approaches to reason about neuroanatomical experiments

Gully A. P. C. Burns, T. A. Russ, T. Ingulfsen, A. M. Khan, E. Hovy
Information Sciences Institute / USC, Marina del Rey CA 90292 [http://bmkeg.isi.edu]

289.1 / FF102

Introduction

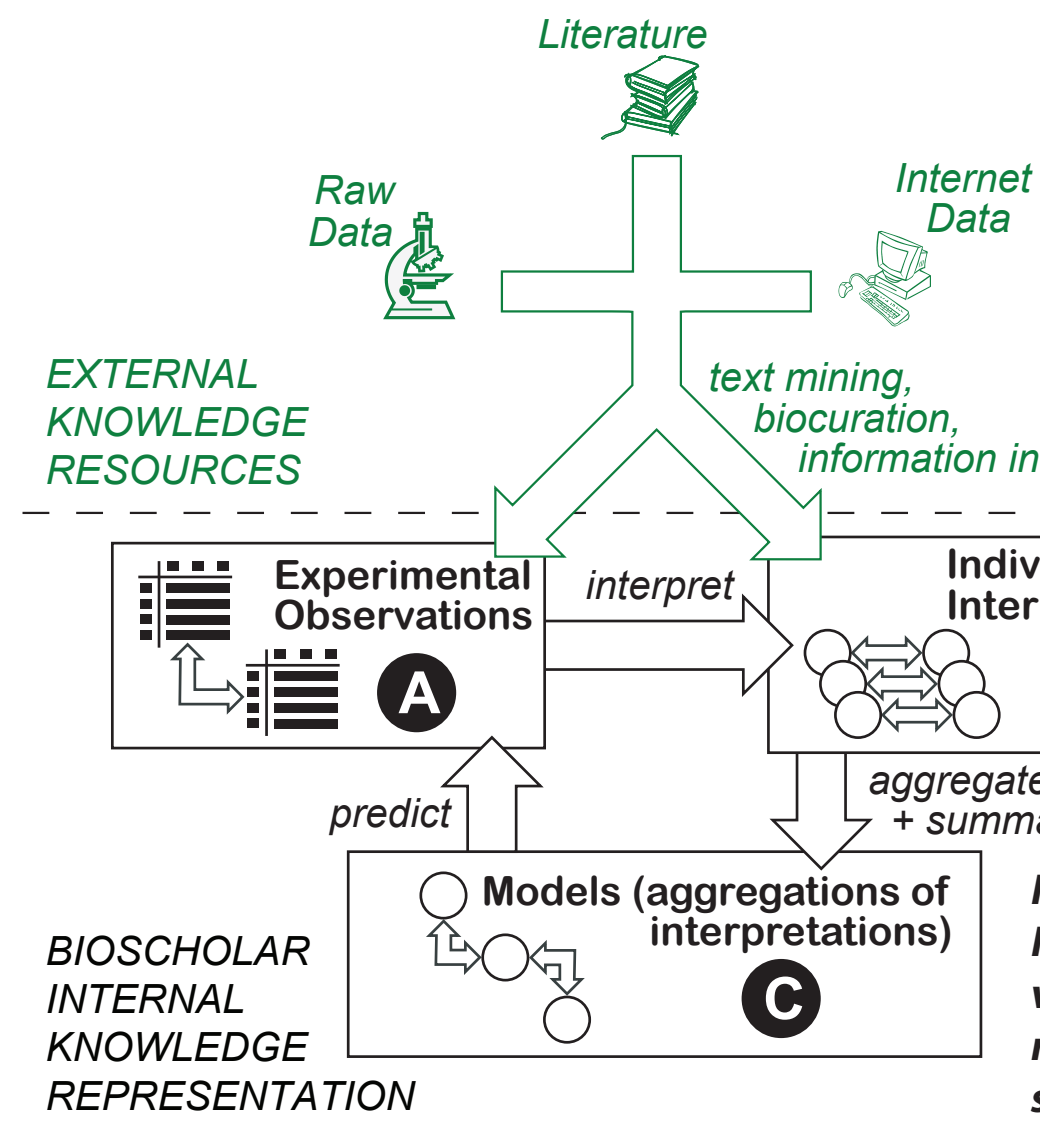


Figure 1: High-level view of modeling strategy.

Here, we describe a general, ontology-based developmen strategy (Figure 1) based on reasoning over experimental observations to generate interpretations that may then be aggregated and summarized into models.

An excellent illustrative example of this approach can be made with *neural connectivity* data (i.e. data from neuroanatomical tract-tracing experiments that permits us to understand neural projection patterns between brain regions). These experiments involve making microinjections of tracers into brain regions of interest. Neurons impinging on the injection site will take up the tracer and transport it via active processes to other parts of its structure, permitting the neurons to be stained and mapped (Figure 2). We wish to build a framework to capture this data and construct a large-scale 'connectivity model'. This framework will subsequently be able to represent data from other types of experiments in the future.

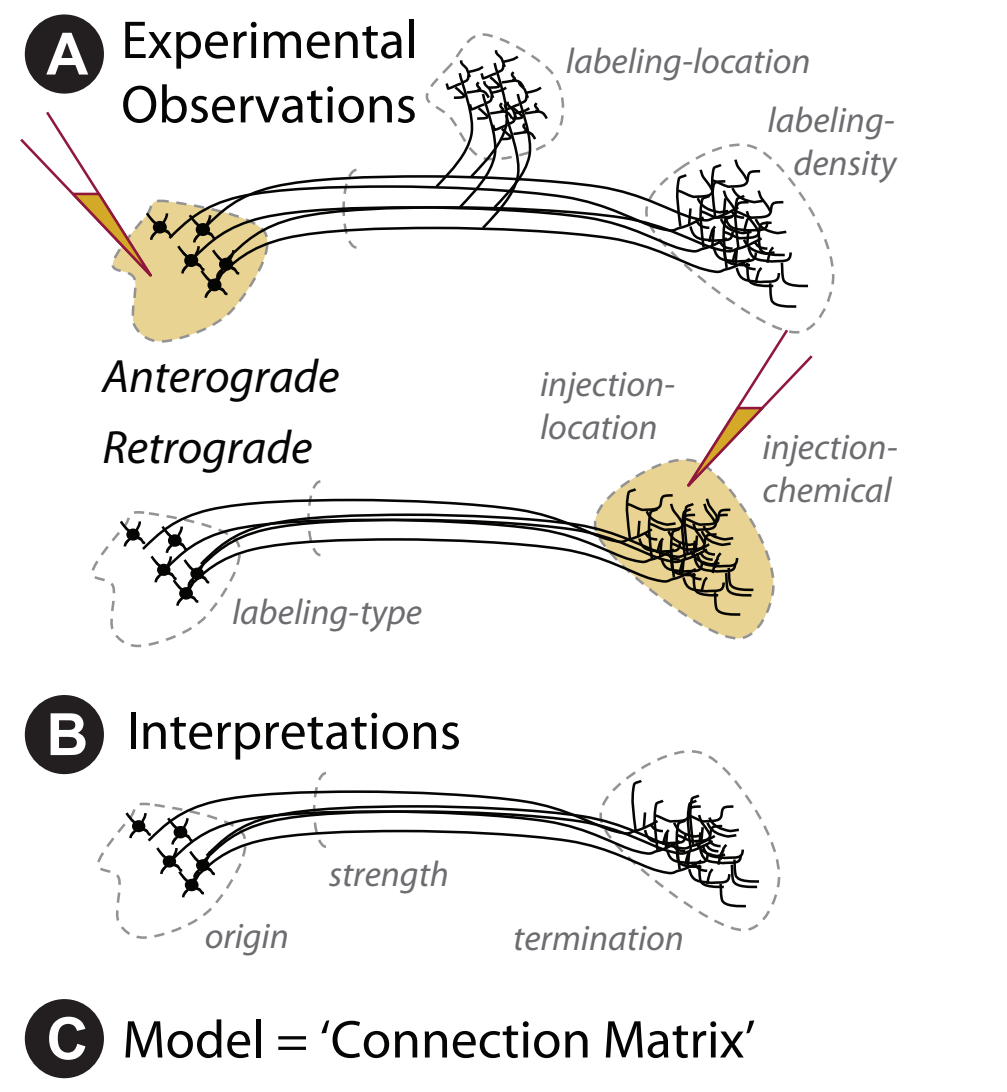


Figure 2: the modeling strategy applied to Neural Connectivity Data.

Knowledge Engineering from Experimental Design ('KEfED')

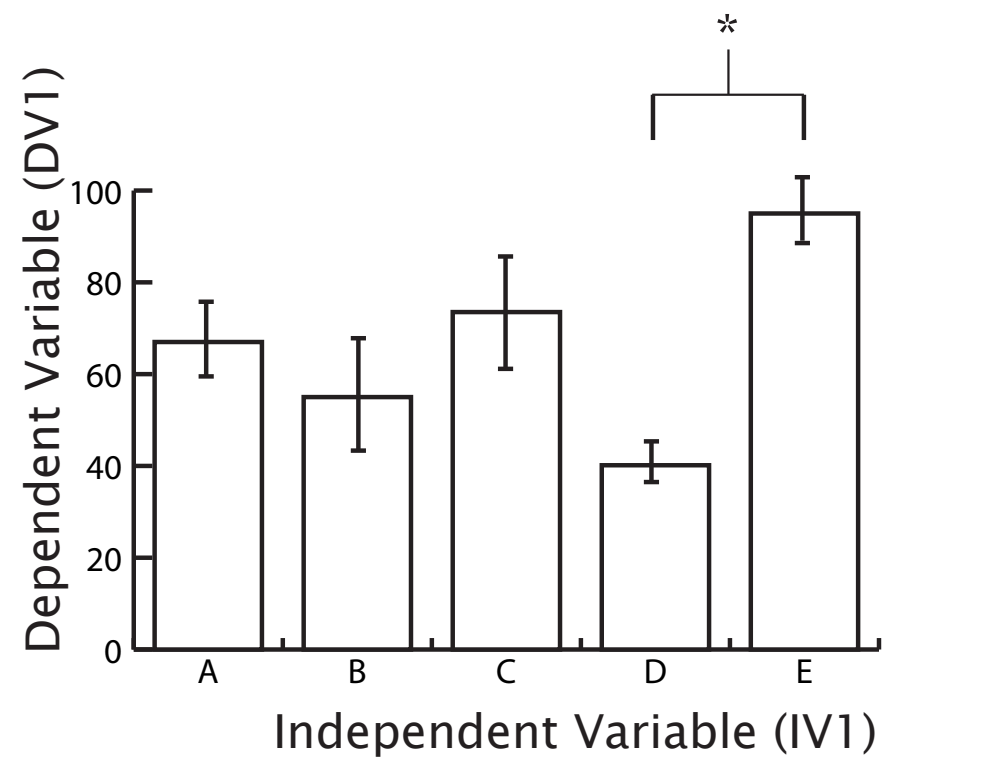


Figure 3: We focus on statistically-significant differences in the value of a dependent variable (a measurement) when changes are made in an independent variable (a constraint).

'Knowledge Engineering from Experimental Design" (KEfED) is a general purpose approach to representing a biomedical experiment to provide a manageable template for that experiment's data. This model is based on the relations between independent (constraints) and dependent variables (measurements) since often, the underlying basis of a scientific 'fact' is a statistically significant effect demonstrated by a difference in the measurement of a dependent variable between cases where an independent variable changes value (Figure 3). We treat each dependent variable as a 'multidimensional vector space' where each axis is denoted by each independent variable (Figure 4). This permits us to represent individual data points, statistical relations between data points, and even statisical correlations. Here we describe preliminary results pertaining to reasoning over data points.

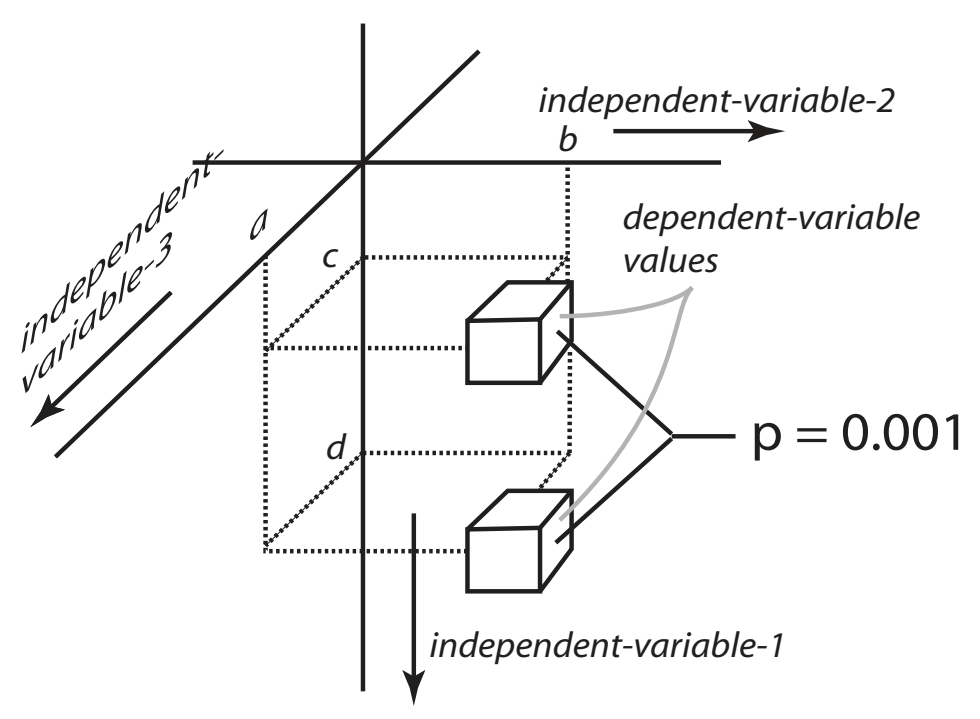
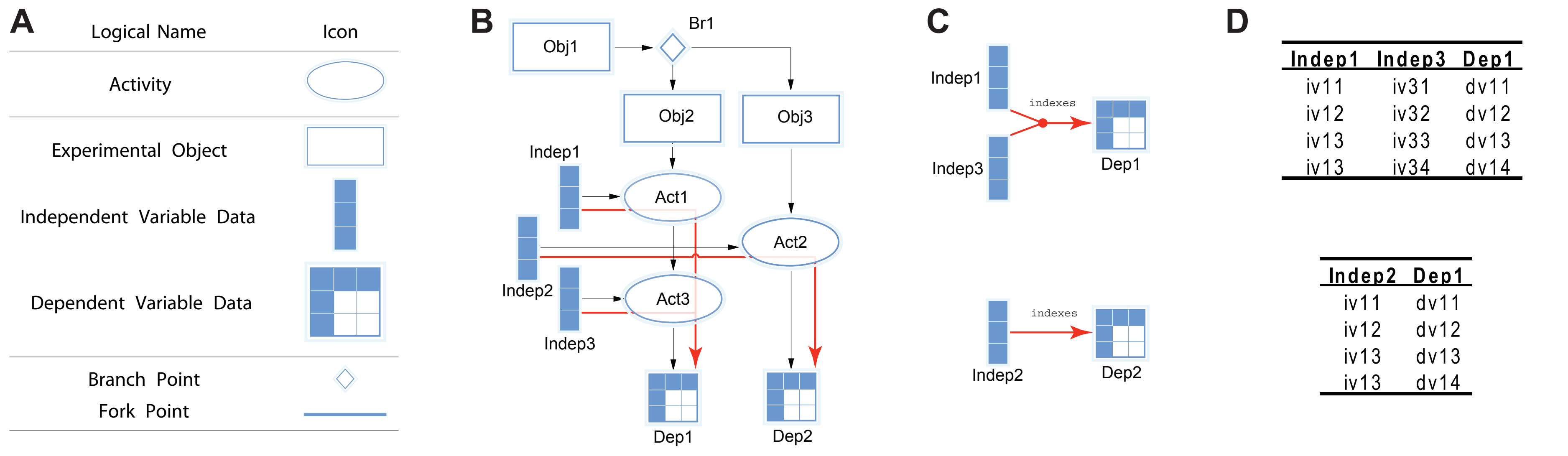


Figure 4: The fundamental organization of the KEfED model. Each dependent variable is a multidimensional array with the independent variables acting as indices for the array.



Indep1	Indep3	Dep1
iv11	iv31	dv11
iv12	iv32	dv12
iv13	iv33	dv13
iv13	iv34	dv14

Indep2	Dep1
iv11	dv11
iv12	dv12
iv13	dv13
iv13	dv14

Figure 5: Components and mechanisms within the KEfED model. (A) Each model is made up of these elements (Activities, Objects , Variables and control flow elements). Note that the semantics of these elements may be complex, and will certainly need to be refined as we instantiate the approach in new systems. (B) A 'fake' example: showing multiple Experimental Objects, Activities and Variables. (C) The dependencies of varaibles in [B] based on pathways through the protocol. (D) Tabulated 'fake' data for the two dependent variables in this example.

The expressivity of the KEfED approach

The KEfED model is potentially significant for the following reasons: (1) It is conceptually simple; (2) It is generally applicable; (3) It is comprehensible to biologists; (4) It supports inference based on experimental findings, rather than intermediate concepts. We have used it to capture the structure of complex experimental designs. Figure 6 shows a neuroendocrinology experiment infusionsjof neurotransmitter are made into a location of the brain and then immunohistochemistr and *in-sit*u hybridization is used to reveal expression of proteins and gene products in target brain regions. The KEfED model may be applied to construct a data repository for such an experiment (see Poster 199.12, Jacobs et al. 2009).

This work is concerned with developing a testable, proof-of-concept for the reasoning required to implement the KEfED model in support of studying neural connectivity. Although this effort is not currently fully compliant with OBO-foundry principles, we are directly collaborating with members of the OBI (Ontology for Biomedical Investigation) consortium firstly to provide standardized terms for experimental elements and to leverage our user interface to assist with ontology curation.We feel that this work is directly complementary to the OBI effort by providing a useful heuristic that naturally limits the depth of representation (we only seek to represent the primary experimental observations necessary to reconstruct valid interpretations from denoted experiments) and provides access of the details of our underlying knowledge representations in terms that are intuitively understandable for biologists.

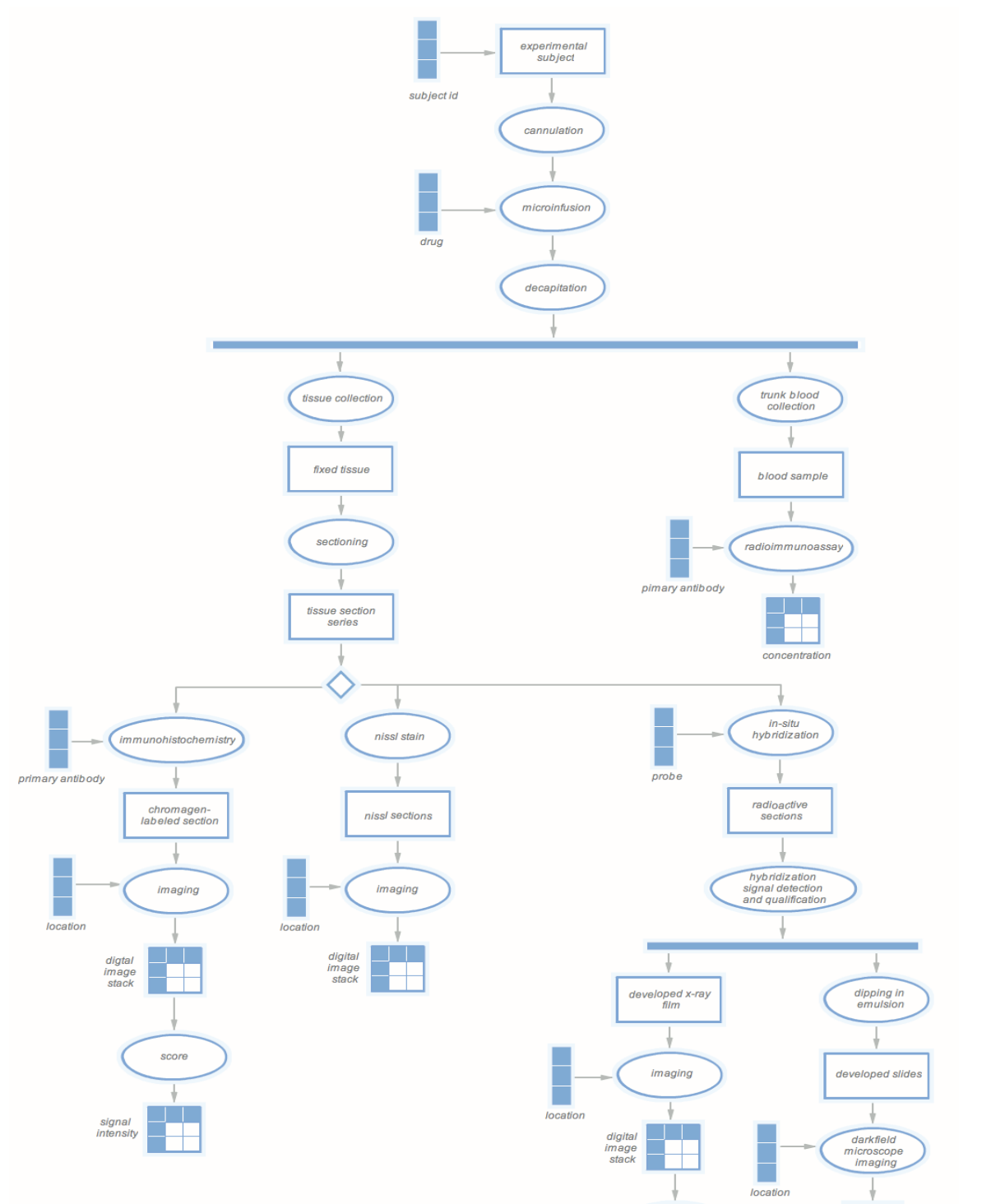


Figure 6: KEfED model for the experimental design for experiment 2 of Khan et al (2007), J Neurosci, 27(27): p. 7344-60.

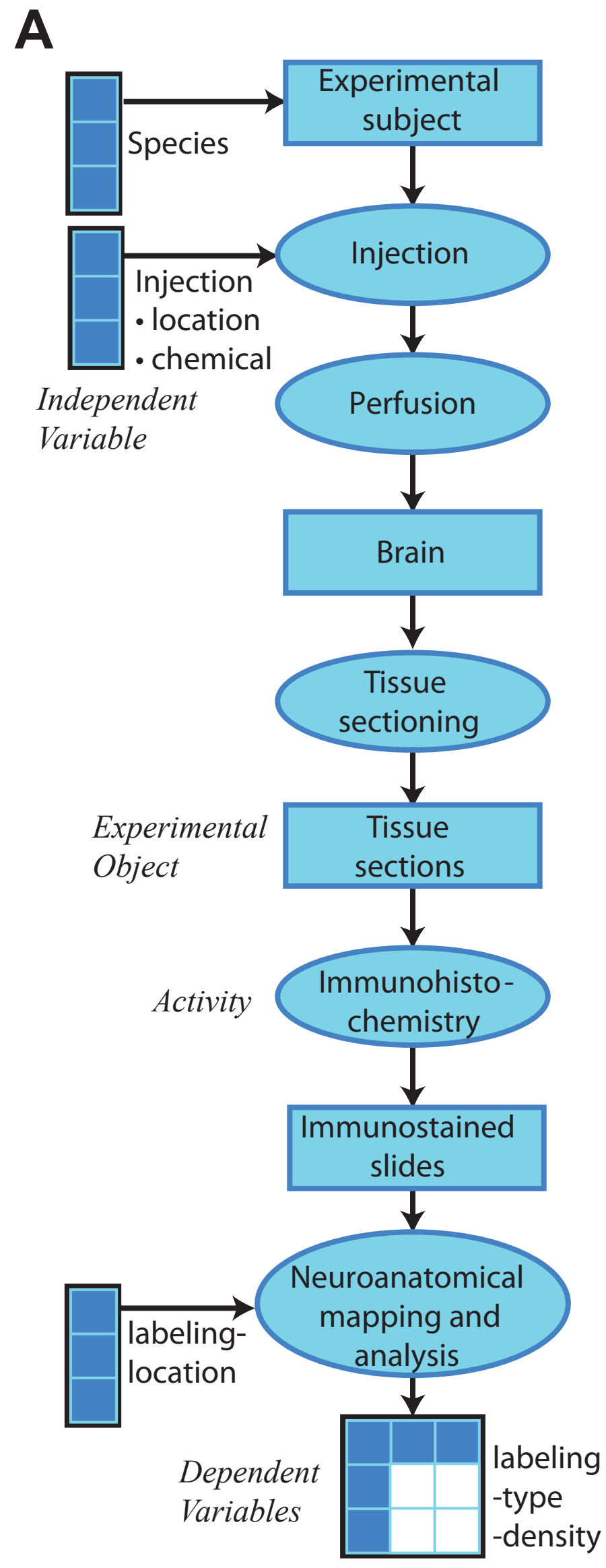
Reasoning with the Powerloom System

We use the PowerLoom [PowerLoom] first-order logic knowledge representation and reasoning system to reason over KEfED models. PowerLoom provides us with a deductive reasoning engine that supports numerical calculations, n-ary relations and closed-world reasoning. It has an integrated query language that allows us to access the information from our encoding of the experimental structures. We use queries and logic inference rules to convert these representations of *experimental observations* into *interpretable statements* pertaining to the information under study. For tract-tracing, we wish to make statements concerning the existence and strength of connections between brain structures.
Table 1 compares OWL and PowerLoom. Expressive power, closed-world reasoning and numeric calculations are key features of PowerLoom that we use for KEfED reasoning.

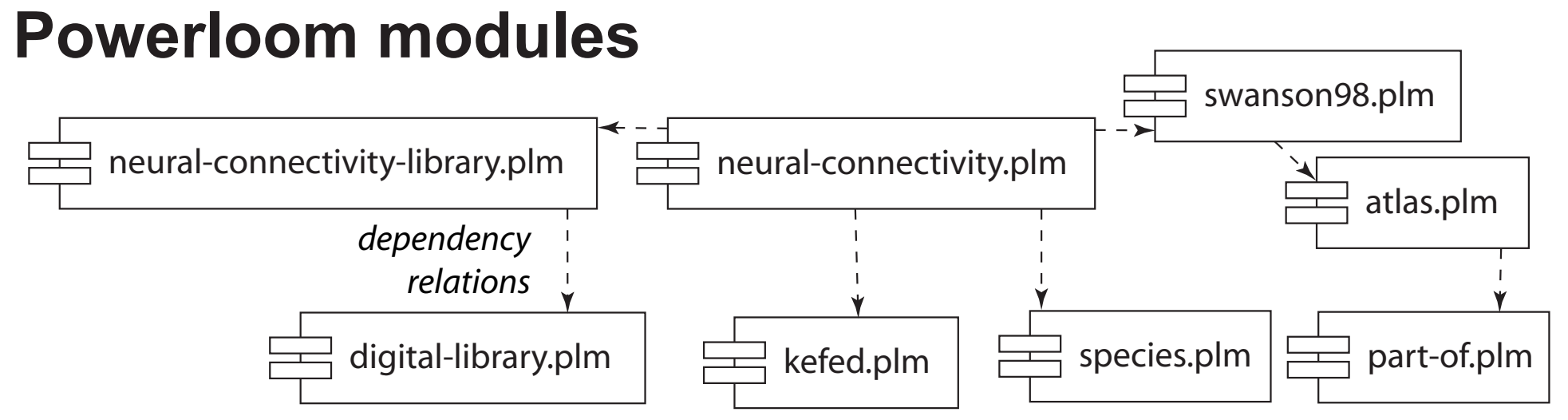
Table 1: Comparison between the Web Ontology Language (OWL) and the PowerLoom first order logic reasoning system.

OWL	PowerLoom
Description Logic	More expressive First Order Logic
Open World	Open and Closed World
Strict Reasoning	Strict and Default Reasoning
Complete Classifier	Limited Classifier
Monotonic Reasoning	Nonmonotonic Reasoning
Rules via SWRL	Built-in Logic Rules
Query via SQWRL	Built-in Query Language
Limited arithmetic via SWRL	Arithmetic
Units Representation only	Units Rep. & Computation
No Cross-property Constraints	Cross-property Constraints
N-ary Relations Reified	Direct N-ary Relation Support
Single Assertion Space	Multiple Assertion Spaces

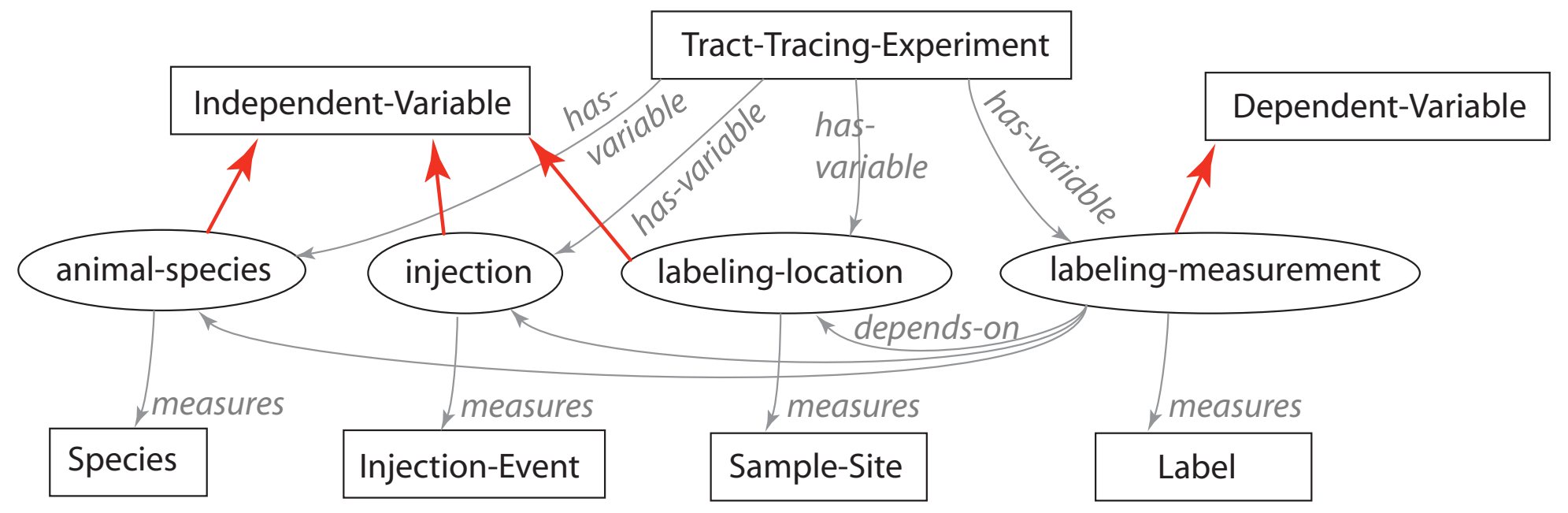
Applying KEfED-based reasoning to neural connectivity



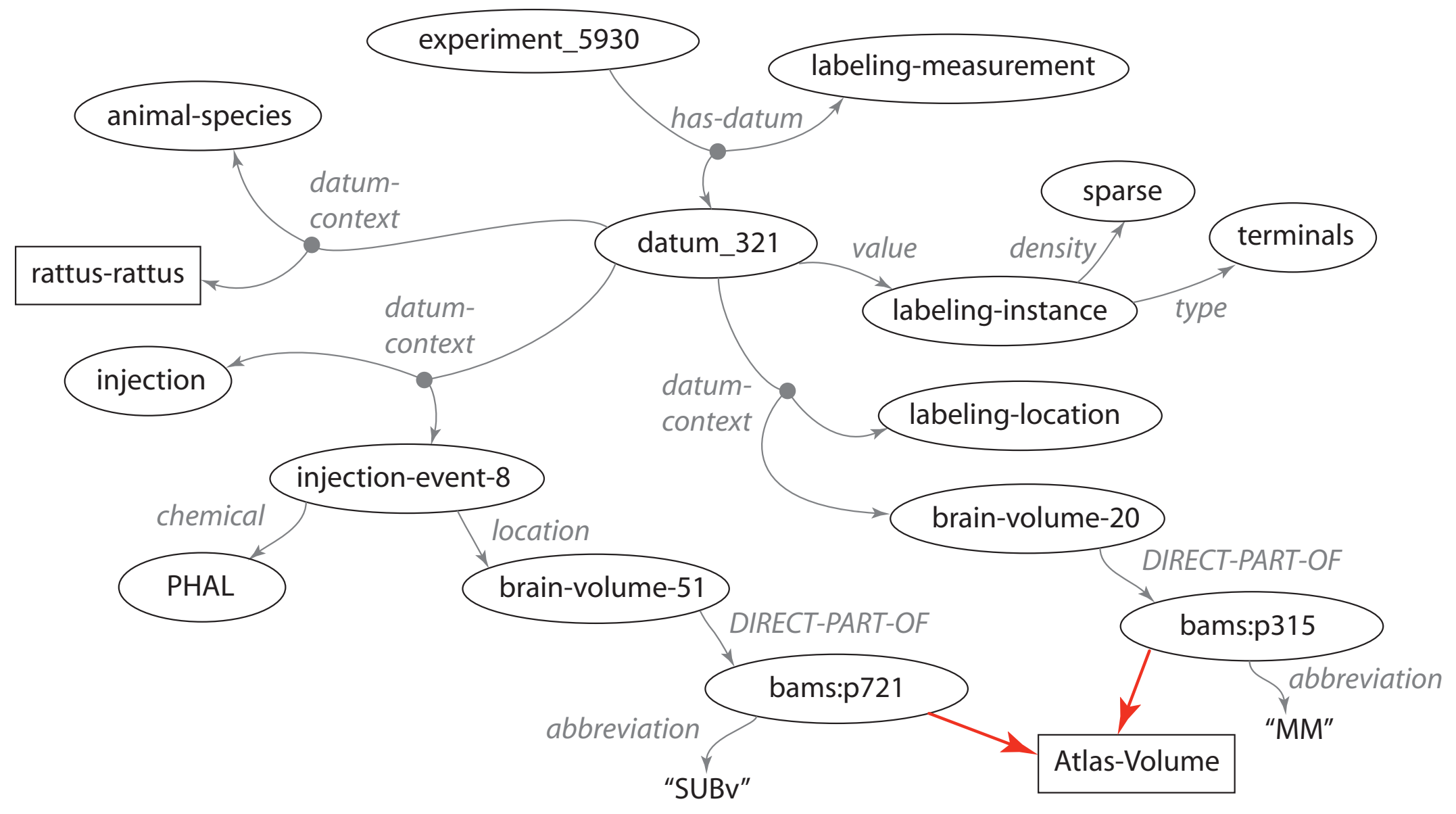
B: Powerloom modules



C: KEfED model for Tract-tracing



D: Assertions for a particular tract-tracing experiment



E: Query: anterograde connections from SUBv to MM

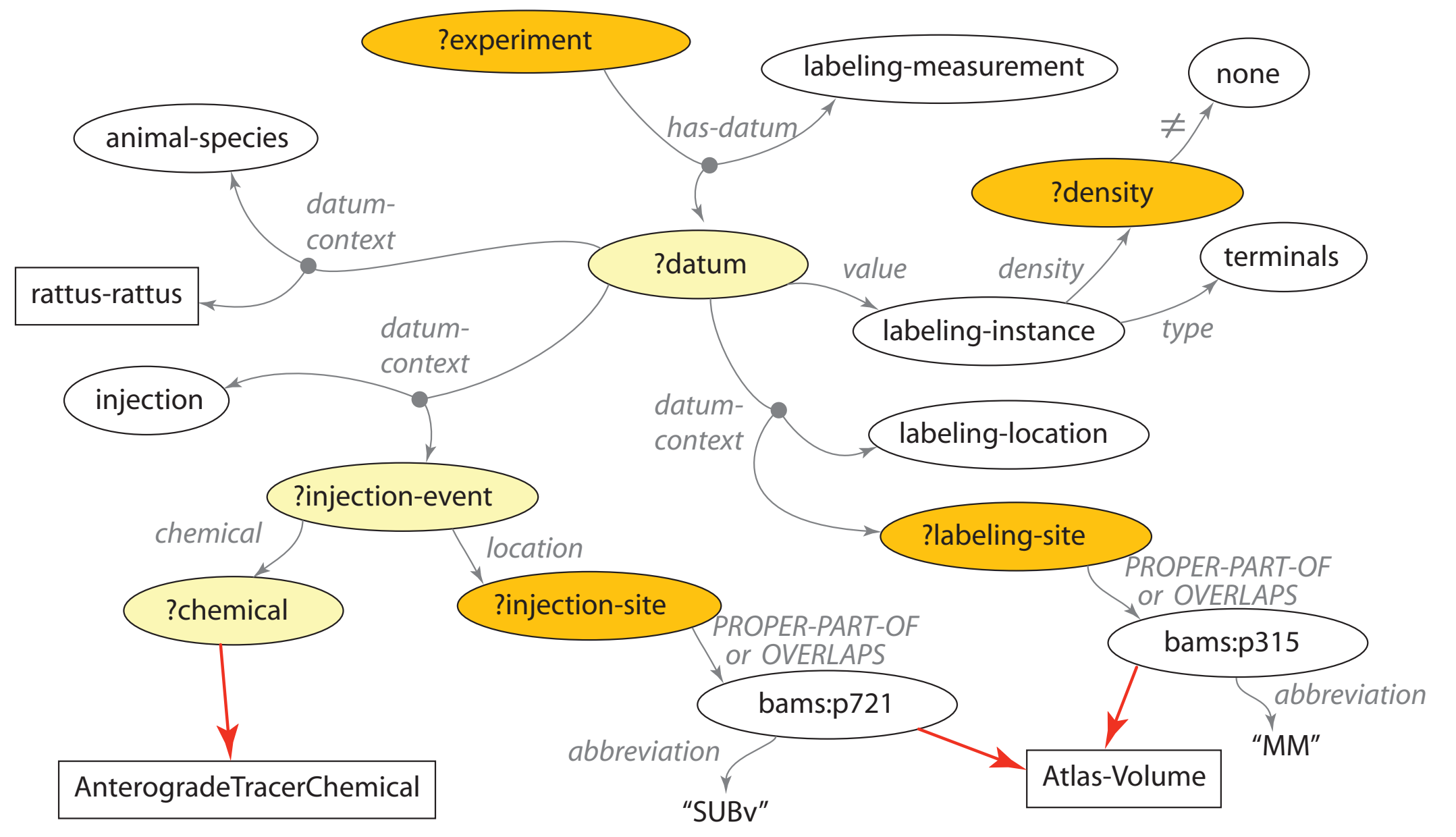


Figure 7:

BioScholar: A web-based system for KEfED models

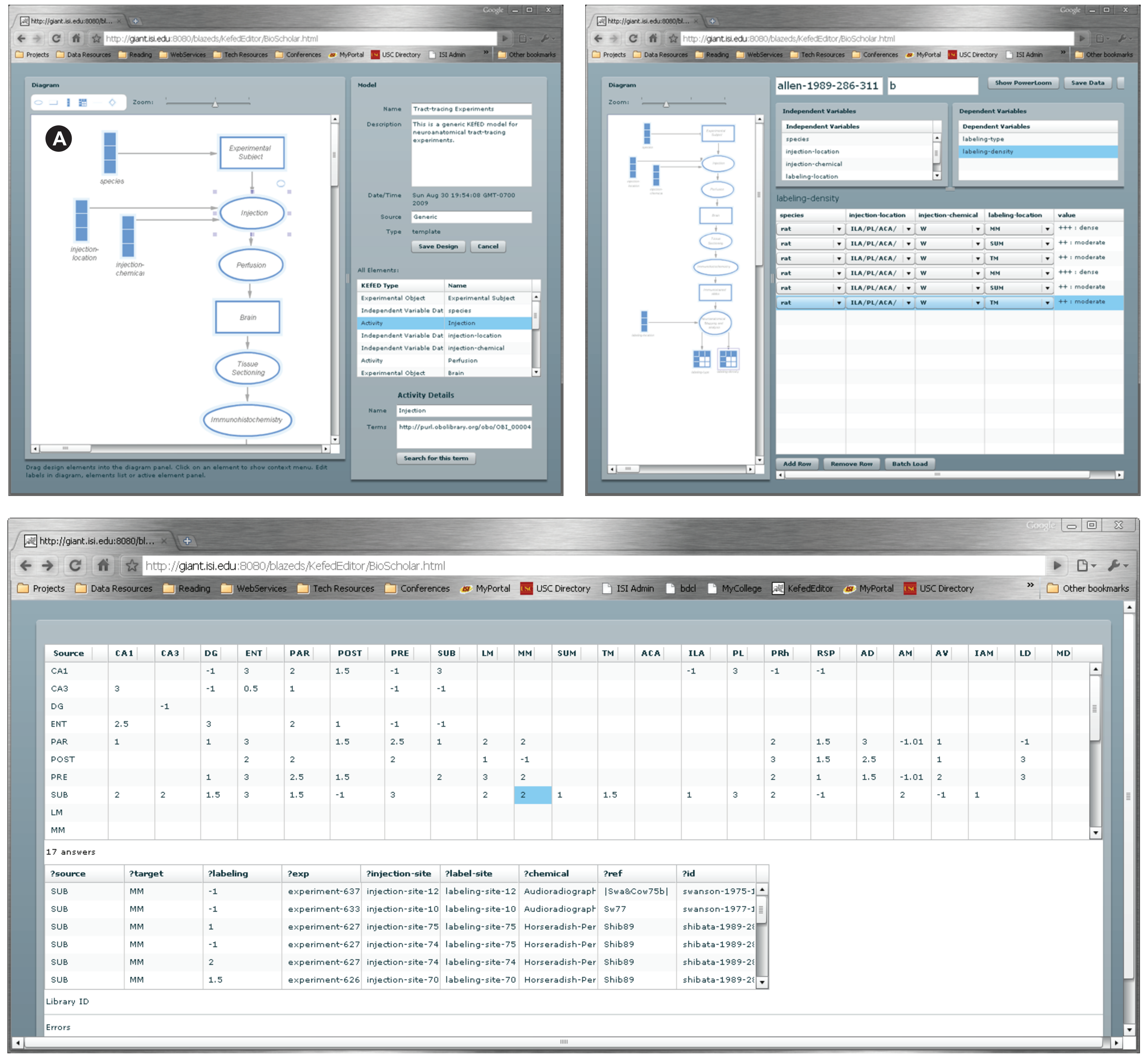


Figure 8-: The BioScholar system (v0.1). This shows the Designer, Spreadsheet and Connection Matrix Components.

BioScholar is currently an early prototype system (shown in Figure 6). We use a graphical editor to build a description of the workflow of an experiment (Figure 6, left). We implemented this using Kap-Lab's 'Diagrammer' control (`http://labs.kapit.fr/display/diagrammer/Diagrammer`) and the Flare Prefuse ActionScript library (`http://flare.prefuse.org`). Links to external ontologies are managed through a control that executes a remote search on the NCBO Bioportal site. Data may be entered into a tabular form (Figure 6, middle) that is automatically derived from the KEfED model of the experiment and then transformed into the logical representation used by the reasoning engine. The reasoning engine is used to generate a connection matrix (Figure 6, right) by querying the system for connection reports from a source to a target region. This query automatically includes 'part-of' reasoning in compiling the results, which are shown in a tabular form after double-clicking an individual matrix cell. Future work involves linking the elements within the representation to community driven ontologies (and providing means to submit new terms to these ontologies), and the development of text mining tools to semi-automate the process of curating articles into the system. Clicking on a connection report in the demo navigates to a marked-up version of the paper where text pertaining to the variables within the model are highlighted.

Future Work

This research is funded by the U.S. National Institutes of Health under grant GM-083871 for the 'BioScholar' project http://bmkeg.isi.edu/. We wish to acknolwedge the programming contributions of Tommy Ingulfsen
for ongoing discussions with many contributors. Special thanks go to Arshad Khan, Alan Watts, Joel Hahn, Larry Swanson, Mihail Bota, Alan Ruttenberg and Hans Chalupsky.