

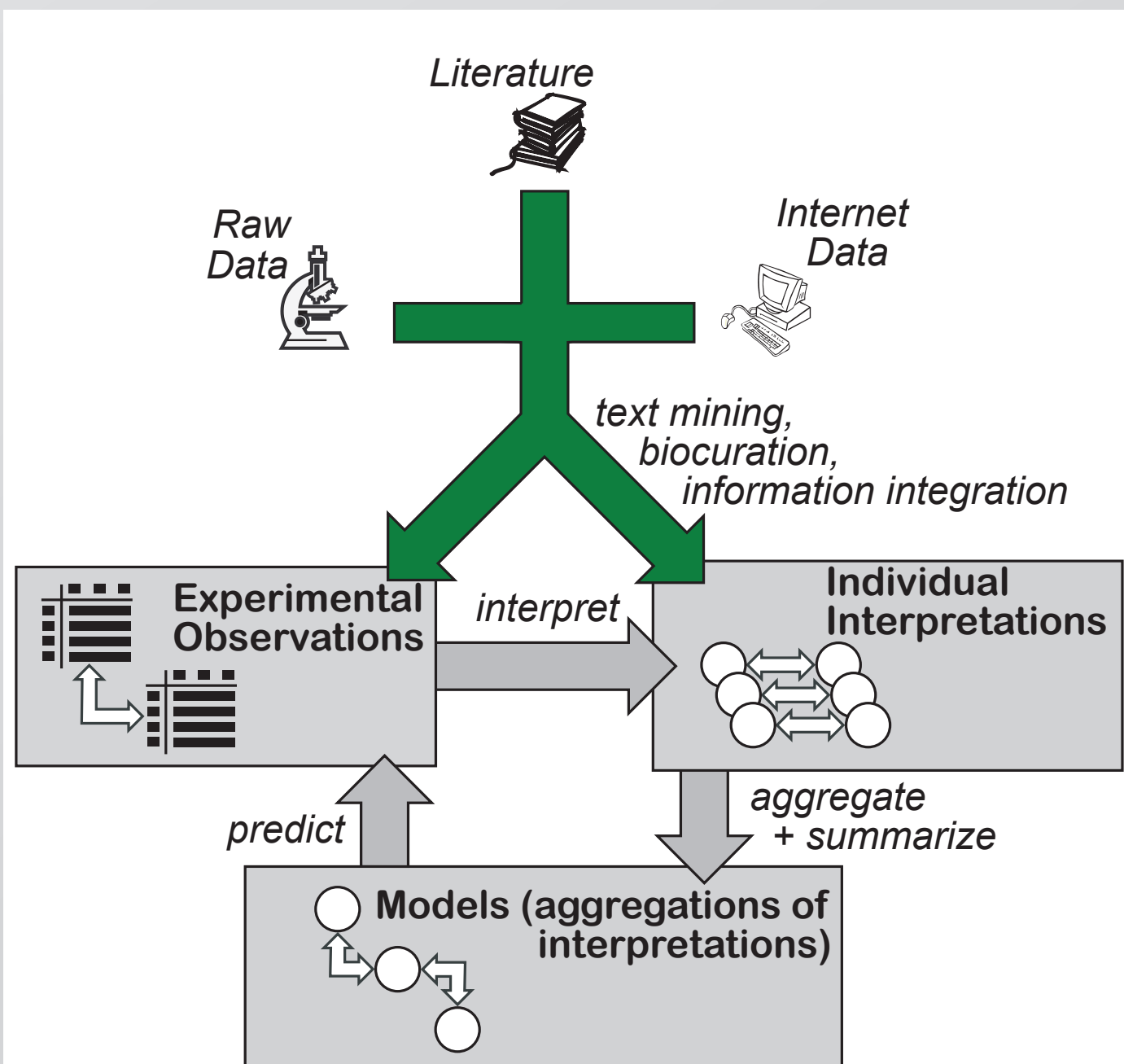
# Biomedical Knowledge Engineering tools based on Experimental Design

## A case study based on neuroanatomical tract-tracing experiments

Gully APC Burns  
Thomas A. Russ

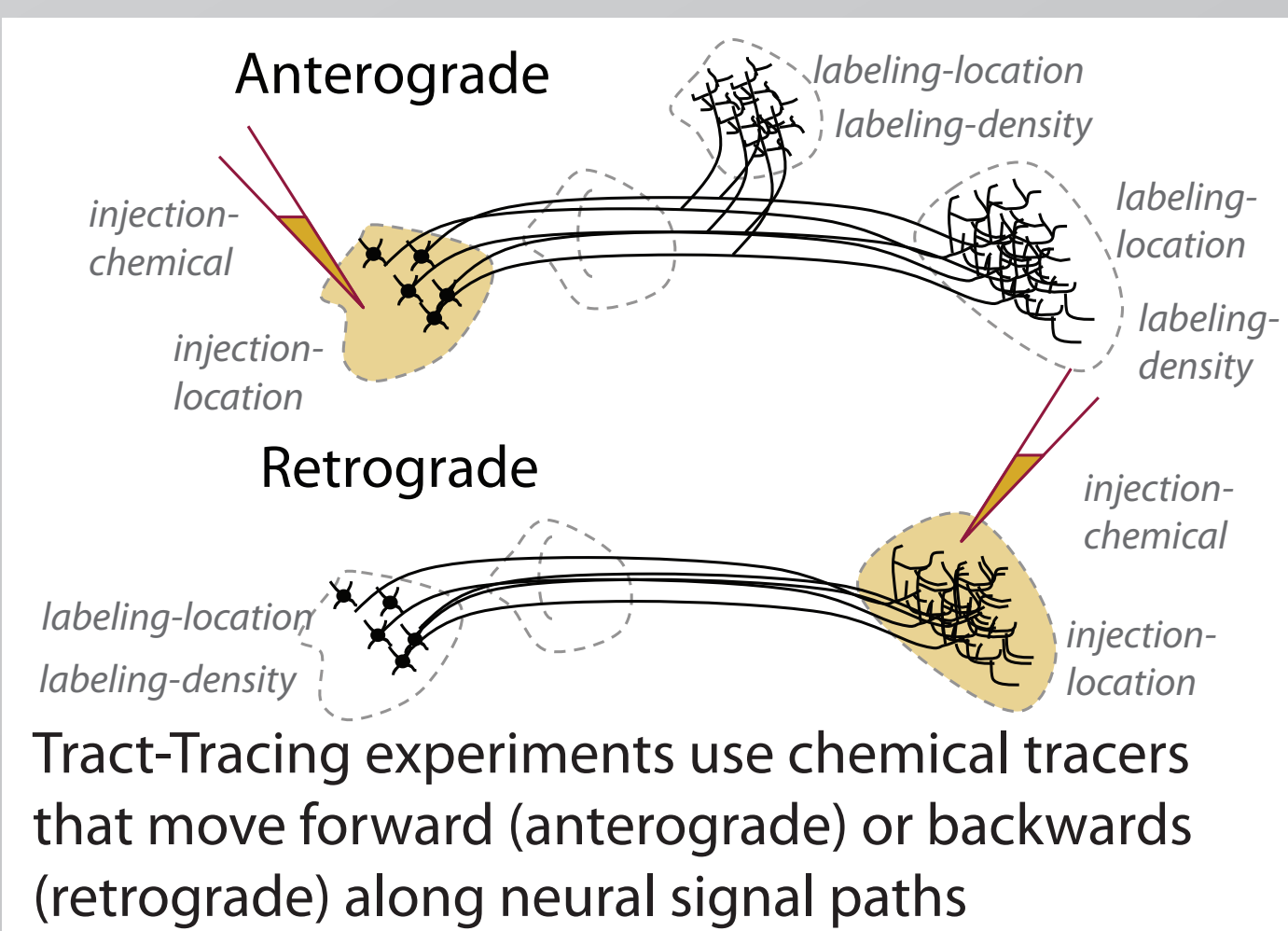
USC Information Sciences Institute  
4676 Admiralty Way  
Marina del Rey, CA 90245, USA

burns@isi.edu  
tar@isi.edu  
<http://bmkeg.isi.edu/>

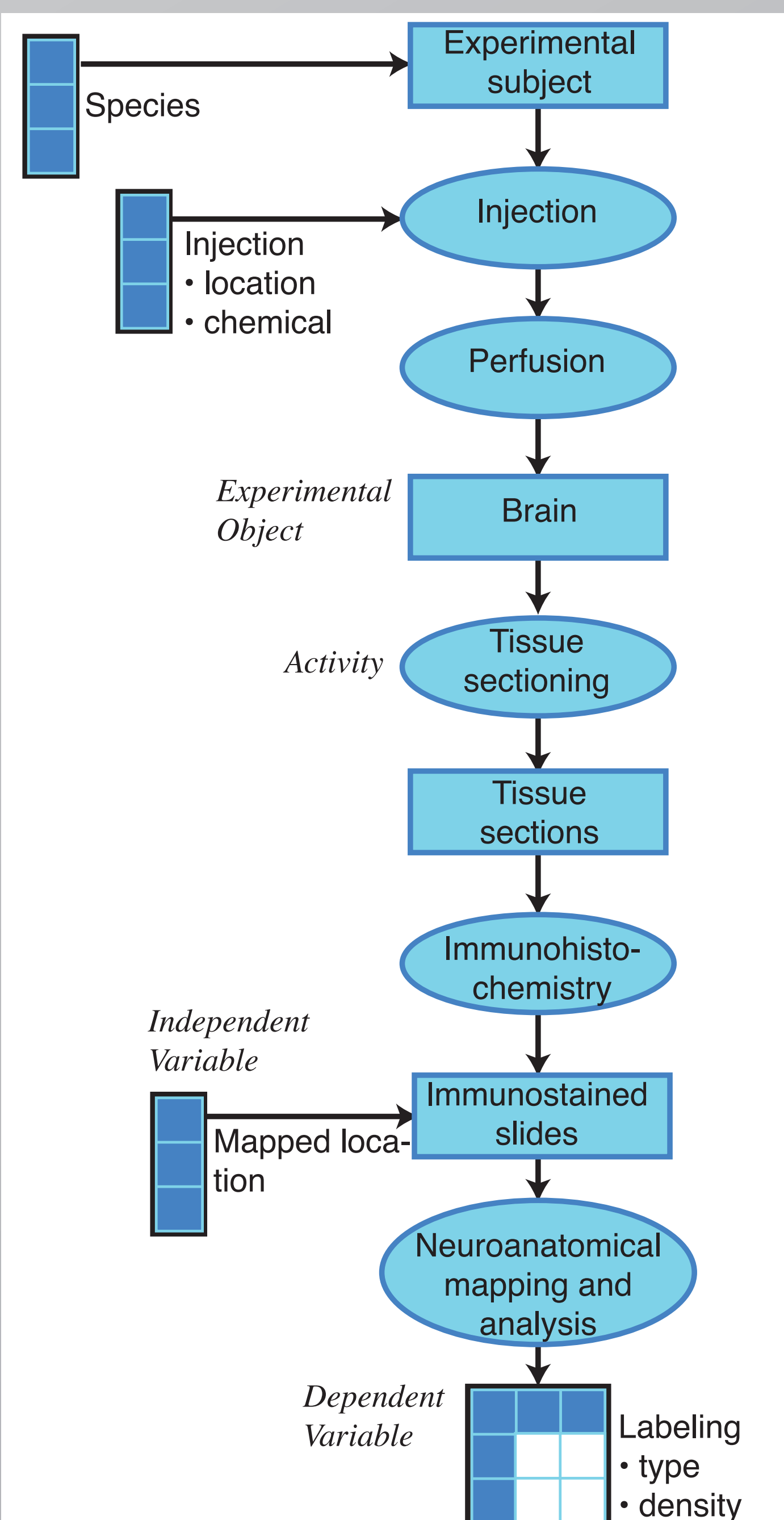


Scientific experimentation involves designing experiments and making observations. Those observations are then interpreted and aggregated into models which explain and predict additional observations.

Our structure for representing experimental protocols and their data is based on this structure.



Tract-Tracing experiments use chemical tracers that move forward (anterograde) or backwards (retrograde) along neural signal paths



KEFED Model for a Tract Tracing Experiment

We are developing a *general-purpose* approach to representing the design of a biomedical experiment in order to provide a manageable template for that experiment's data. We call this approach "Knowledge Engineering from Experimental Design" or (KEFED) [kefed]. The use of the KEFED model gives us the ability to impose a formal and well-grounded structure on the information contained in scientific articles, based on relationships between dependent and independent variables that make up a scientific experiment. This structure then allows us to add value to the information contained in the articles by performing directed information retrieval, adding basic forms of reasoning based on additional information such as anatomical atlases and taxonomies from external ontologies. We present a graphical interface for constructing KEFED models and a first-order logic reasoning system that performs inference over such models.

A key organizing principle for our data storage is the relationship between dependent and independent variables in an experiment.

**Independent Variables** are variables whose value is set as part of the experimental design. They are independent because their value does not depend on any other variables in the experiment.

**Dependent Variables** measure experimental results in the context of a set of values for independent variables. The values depend (are potentially influenced by) the values of the independent variables in the experimental design. We use the experimental design to allow us to trace the dependencies of the variables.

The values taken on by independent variables belong to a particular domain. This domain and the values in it can be linked to external ontologies to specify the meaning of the domain and values.

### Tract-tracing Experiments

Tract-tracing experiments involve making microinjections of tracer chemicals into carefully defined locations of the brains of laboratory animals. Parts of neurons in those locations take up minute quantities of tracer chemical to be transported by the neurons' internal molecular machinery along their long axonal processes. The animal is then killed and its brain examined to locate the tracer, thus revealing the presence of connections between spatially separate parts of the brain. Most of our understanding of the internal wiring of the brains of mammals has been discovered through the use of this technique [Swanson].

In our demo, we show how the data from such experiments can be used with a reasoning engine to show support for connections.

### Web-based KEFED Editor

KEFED uses a graphical editor to build a description of the work flow of an experiment. This produces a graphical model, such as the model of the tract-tracing experiment at left. We use the work flow to identify the context for data values. The data is entered in a tabular form automatically derived from the KEFED model of the experiment type and then transformed into the logical representation used by the reasoning engine.

We implemented the grapher using Kap-Lab's Flex Diagrammer (<http://labs.kapit.fr/display/diagrammer/Diagrammer>) and use a graph representation from the Flare Prefuse ActionsScript library (<http://flare.prefuse.org/>) so we can use their graph-traversal and shortest-path algorithms.

### Reasoning over Tract-tracing Data

We perform our reasoning using the PowerLoom [PowerLoom] first-order logic knowledge representation and reasoning system. PowerLoom provides us with a deductive reasoning engine that supports numerical calculations, n-ary relations and closed-world reasoning. It has a query language that allows us to access the information from our encoding of the experimental structures.

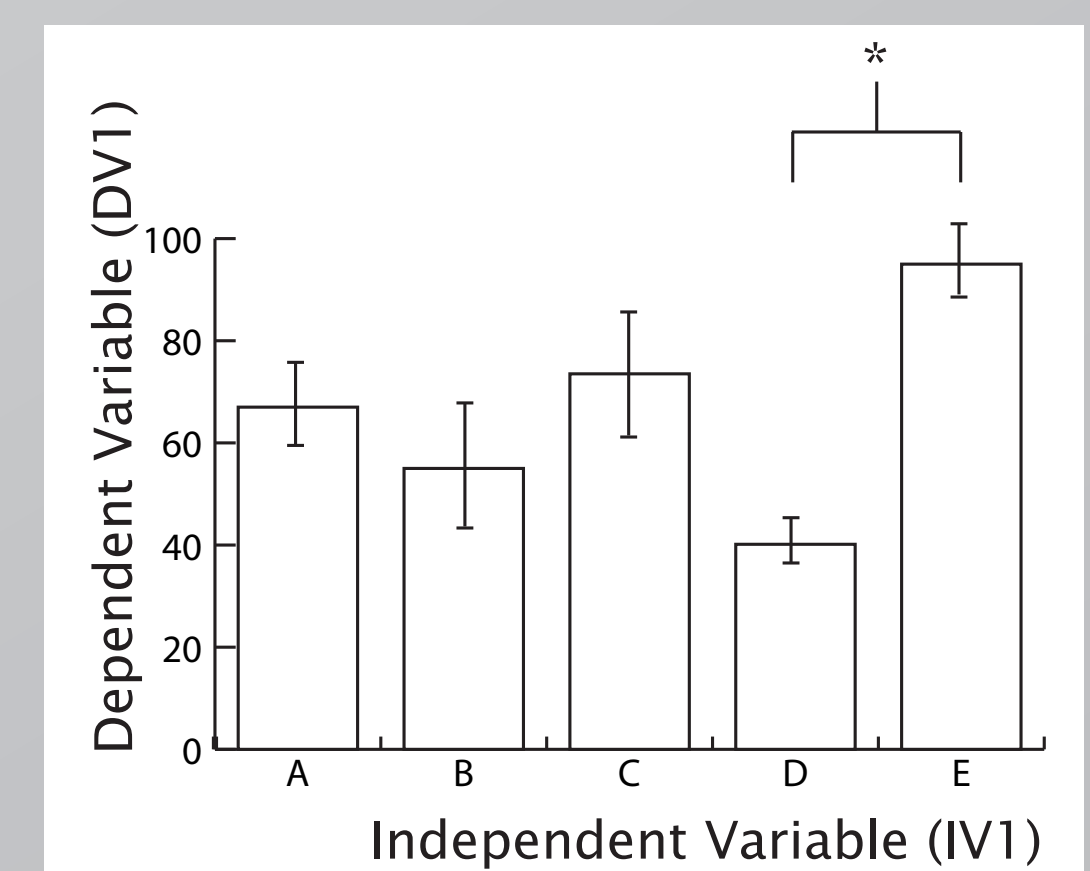
We import the basic geometric relationships from the Swanson brain atlas [BAMS] into PowerLoom and use it to reason about geometric relationships between different brain regions.

### Future Work

Our future plans include applying text-mining tools to help process documents and speed the data acquisition phase. These tools will then be packaged into a system called *BioScholar*.

- [BAMS] M. Bota and L. Swanson. The brain architecture management system. <http://brancusi.usc.edu/bkms/> 2009.
- [kefed] G. Burns, E. Hovy, and T. Ingulfsen. Biomedical knowledge engineering approaches driven by processing the primary experimental literature, in *Frontiers in Neuroinformatics. Conference Abstract: Neuroinformatics 2008*.
- [PowerLoom] PowerLoom. PowerLoom knowledge representation & reasoning system. <http://www.isi.edu/isd/LOOM/PowerLoom/> 2009.
- [Swanson] L. W. Swanson. *Brain Architecture, Understanding the Basic Plan*, Oxford University Press, Oxford, 2003

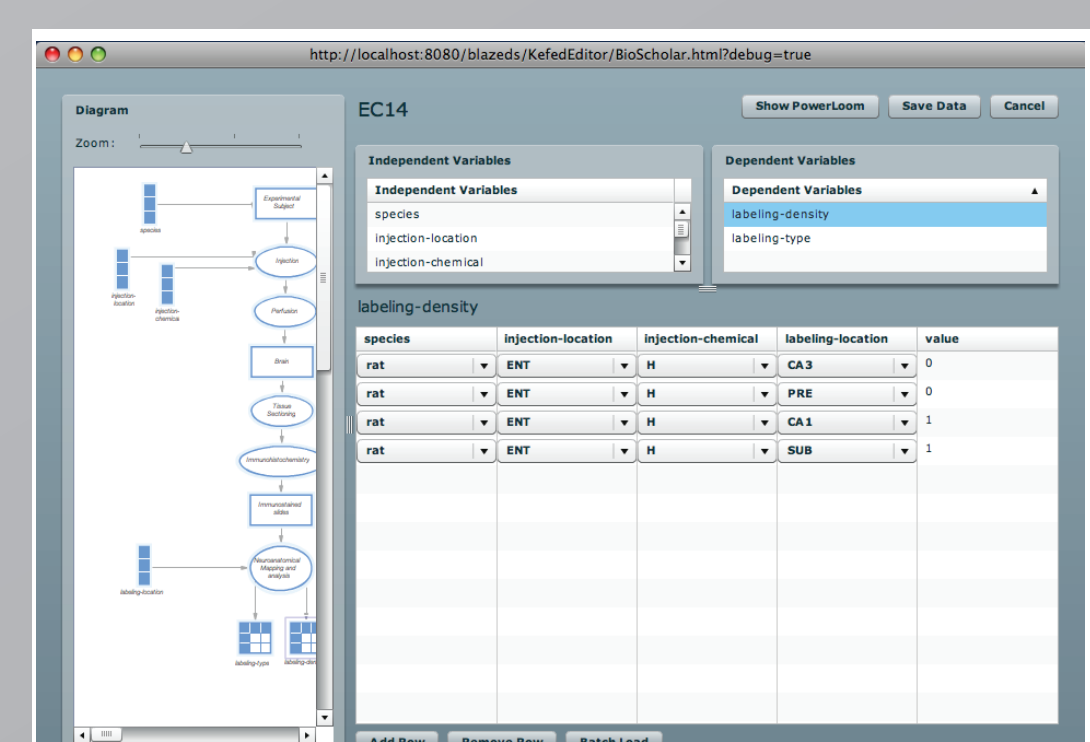
This research is funded by the U.S. National Institutes of Health under grant GM-083871 for the 'BioScholar' project <http://bmkeg.isi.edu/>. We wish to acknowledge the programming contributions of Tommy Ingulfsen to the BioScholar project. The ideas in this work are the culmination of many discussions. Special thanks go to Arshad Khan, Alan Watts, Joel Hahn, Larry Swanson, Alan Ruttenberg, Ed Hovy and Hans Chalupsky. Additional thanks to Clay Morrison and Judea Pearl for on-going discussions concerning the use of the KEFED model in causal reasoning.



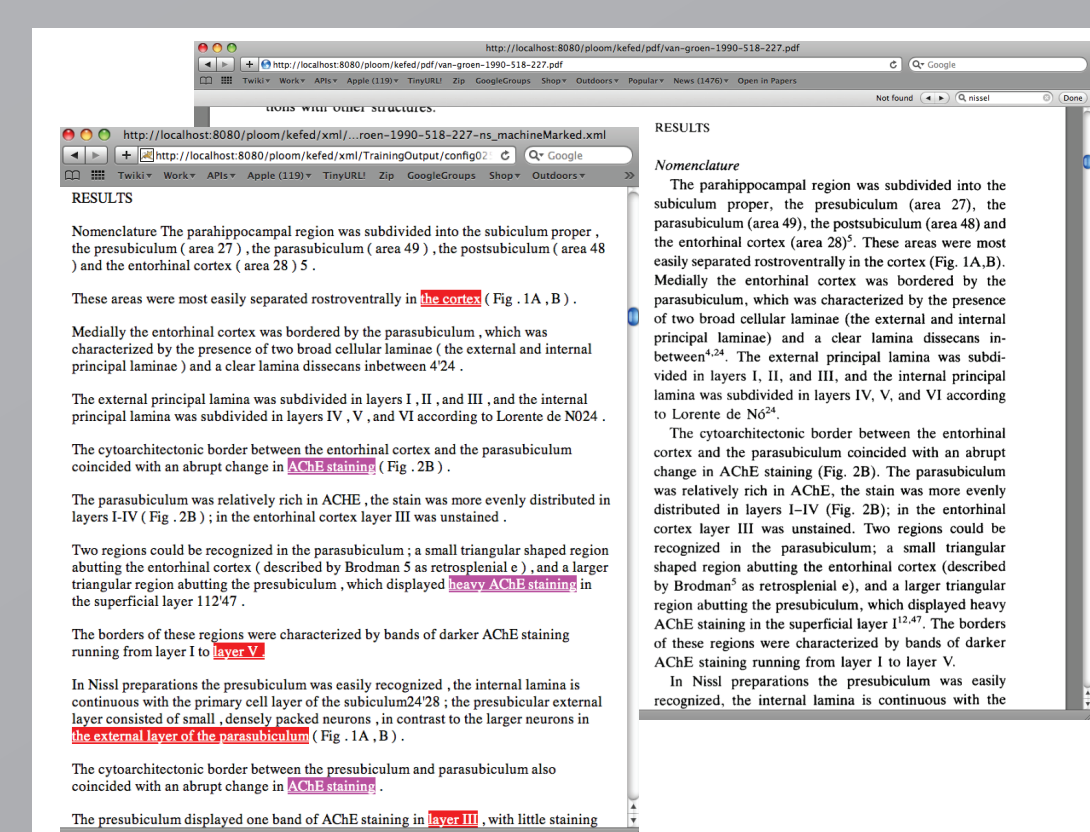
Result graph showing the relation between dependent and independent variables. Experiments seek to find a *significant difference* between values.

The screenshot shows a 'Tract Tracing' table with columns for 'Source' and 'Target'. Below it is a table of '5 sources' with columns for 'Source', 'Target', 'Labeling', 'Inj', 'Injection-chem', 'Labeling-chem', 'Injection-loc', 'Labeling-loc', 'Inj-chem', and 'Labeling-chem'. The table lists various experimental conditions and their corresponding results.

Connection matrix and supporting experiments for one of the cells. We use PowerLoom for reasoning about the underlying data.



KEFED editor data entry form for the labeling-density variable in a tract-tracing experiment. The form is constructed from the variable dependency information from the design.



Future Work: Applying text-mining tools to help process documents