# Whither History of Science in Discovery Informatics?

2012 NSF Workshop on Discovery Informatics

Will Bridewell
Stanford Center for Biomedical Informatics Research
Stanford University

Position papers are strange creatures that encourage you to trot out all your old cliches. For instance, I might talk about the innovative treatments of data, carving them up in a variety of ways to find patterns that explain anomalies and rare phenomena (Bridewell, 2004). Better still, I could talk about the paucity of useful scientific data, a taboo topic in the age of microarrays and large hadron colliders, and the ensuing need to develop approaches that can infer explanations from a handful of time series (Bridewell et al., 2008). I could also chat about the possibilities of learning from underdetermined systems using a combination of systematic search through a solution space and the modern tools of machine learning (Bridewell & Todorovski, 2010). (Imagine systems that could learn from their mistakes instead of repeating them on each restart!) Instead, I think I will take a new position, one that came to mind earlier today and was developed with a colleague while waiting for a meeting to start.

Before I begin, I should say that I am most definitely unsure what Discovery Informatics is. More specifically, I have a hard time distinguishing it from Discovery Science, Science Informatics, eScience, Computational Scientific Discovery, and informatics in general, which at least in the worlds of biology and medicine has more than a passing connection with knowledge discovery. So, I apologize in advance if my position is offensive, out of line, or simply irrelevant. We must all work within our own limitations.

Although many computationally informed efforts concerned with understanding, automating, and improving the scientific enterprise have pilfered the philosophy of science for ideas and motivation, few have ventured into the realm of the history of discovery. This embarrassing want of attention is remarkable since investigations, both cursory and detailed, of historical patterns of scientific thought heavily influenced the development of normative and descriptive theories of science. These theories in turn influenced all major attempts to develop computational discovery systems.

But, you might protest, historical studies are just stories, contextually dependent tales about singular journeys that carry the scientist from rambunctious upstart to world renowned genius. Without question, one independent story is an anecdote, and as we all know, the plural of "anecdote" is not "data," except in cases like this one, when it is. Stepping back, Frederic L. Holmes (2004), one of the foremost historians of science, viewed scientific careers through the metaphor of an investigative pathway, the trail the researcher forges throughout her life in pursuit of knowledge. By exploring these pathways at differing levels of detail, he began to identify patterns and strategies of discovery. A methodology that we can now replicate on the large scale.

Today, we have at hand a wealth of electronic documents in the form of scientific reports, complete with authors, abstracts, and bibliographies. Although individually the documents are formal and somewhat post hoc descriptions of knowledge discovery, they collectively trace each scientist's informal path from earlier to later discoveries within the context of the larger scientific community. Superficial analysis of these materials lets us create from any single reported discovery the long history of influences leading to that point. By adapting existing informatics techniques we can track prior publications by the authors and their research laboratories and identify the most relevant findings from outside the group.

To process the documents now organized into a branching timeline, we can develop methods for extracting conjectures and findings from research reports, a valuable effort in its own right. From these data, or collection of anecdotes if you are so inclined, systems for abductive inference (Bridewell & Langley, 2011) can infer discovery narratives that color the connections between documents. Of course, due to the frequent collaboration of scientists, the inference system would need to be socially aware, capable of tracking the beliefs and goals of individual actors throughout their careers (Bridewell & Isaac, 2011). The result of this effort would be a collection of noisy and potentially incomprehensible investigative pathways, but by grouping them together by scientific disciplines, inductive methods may reveal common research strategies. These narratives would lay bare the processes and patterns of discovery.

## References


Bridewell, W., & Isaac, A. (2011). Recognizing deception: a model of dynamic belief attribution. *Advances in Cognitive Systems: Papers from the 2011 AAAI Fall Symposium*, 50–57.

Bridewell, W., & Langley, P. (2011). A computational account of everyday abductive inference. *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society*, 2289–2294.

Bridewell, W., & Todorovski, L. (2010). The induction and transfer of declarative bias. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, 401–406.

Bridewell, W., Langley, P., Todorovski, L., & Džeroski, S. (2008). Inductive process modeling. *Machine Learning*, *71*, 1–32.

Bridewell, W. (2004). *Science as an Anomaly-Driven Enterprise: A Computational Approach to Generating Acceptable Theory Revisions in the Face of Anomalous Data*. PhD Thesis, Computer Science Department, University of Pittsburgh.

Holmes, F.L. (2004). *Investigative Pathways: Patterns and Stages in the Careers of Experimental Scientists*. New Haven, CT: Yale University Press.