# Justice may be blind: Science is all eyes.

**Steve Sawyer, iSchool, Syracuse University, Syracuse NY, USA**

Prepared for the Discovery Informatics Workshop, Arlington, VA, 2-3 February, 2012

Science is a social activity and we are – as Paul Feyerabend (1975) said so well – "practitioners of a strange craft." My perspective as a social scientist and a socio-technical scholar is to see discovery informatics as an emerging set of social practices enmeshed into an emerging set of science practices, tied into an emerging set of new technological opportunities. Huzzah for this! The wide-open spaces of the future draw us forward. This is framed, however, by the time-worn walking-path trajectories that brought us to this new start.

So, while discovery informatics may be new, it will still be science and we have lots of guidance on what this might look like. Karin Knorr-Cetina (1999) (and many others) makes clear science is bound up in a set of social norms, incentive systems, institutional structures and webs of funding and prestige. Bruno Latour and Steven Woolgar (1979) (and many others) show us how science is framed by beliefs about the nature of what is a fact, what are the means of knowing facts, and how these facts are discovered, decided, translated and communicated. Thomas Hughes (1983) (and many others) makes clear the social constructions of our science both shape, and are shaped by, the technologies and methods we use to reinforce what it is we call science[1].

We know that scholarly disciplines provide us some means to make sense of all this. This is why doctoral training is so important – we need to instill in students a set of practices and beliefs: we need to pass on to them discipline. Inter-disciplinary discourse provides us some means to revel in our differences and find similarities. New ideas provide us sustenance to achieve: the future runs through us. Here we are!

Evidence of what is possible through discovery informatics will flow through our workshop. Many of these possibilities will impress us. The three points I make are mundane in comparison. The suggestion which follows builds from these: the goal is to frame discovery as a local, but increasingly connected, activity.

**Mundane Point 1: The social practices of discovery informatics will look much like the social practices of the scientific practices you currently observe.** The possibilities of discovery informatics will be realized through the slow-changing world of tenure at universities, reduced federal funding and its constrained alternatives, and the discovery-winner symbiosis of research stars and the enthusiasm with bigness. Large data sets and powerful algorithms *might* lead to where a few super-scientists, at a few super places, can become the locus of science. These few places will generate the bulk of new discoveries, attract the lion's share of funding, and host the greatest volume of data. How, then, will the rest of the scientists succeed? What will be the career paths, rewards and opportunities for the rest of us? Should we place science into the hands of the few(er) by making it even more difficult to participate at the leading edge of science? Certainly those who support open access to data worry about this – the trend is visible. Discovery informatics *might* also be advanced as a set of tools and techniques that allow more people to pursue the scientific frontier. If so, resources, reward structures, and training will need to draw scientists to this approach: current resource allocations, reward structures and training do not.

**Mundane Point 2: New approaches to doing science will face stiff competition from those who have mastered the current practices.** Experts do not go away quietly. Imagine a conversation where discovery informatics is advanced as a better way to develop theory. Leading contemporary theorists will do more than scoff: they will actively discourage such heresy. And, they may be right: what insights are used to develop

[1] See also M. Lamont (2010) *How Professors Think: Inside the Curious World of Academic Judgment*, Harvard Univ. Press: MA.

algorithms of discovery? What guides collection of what forms of data? The conversation could be made worse by discovery informatics proponents questioning the need for theory. Why fight this battle? Discovery informatics is a possibility, a confluence of people's paths. Imagine a second conversation, one that begins with what can we do with discovery to leverage our current knowledge? How can we use what we know to direct discovery, to advance our comparative analysis, to situate discovery informatics as a complement to current ways of advancing knowledge? The normative social pressure of disciplines is real and powerful. Play nice: good ideas rarely survive bad introductions.

**Mundane Point 3: All new technologies have intended and unintended consequences. These consequences will be both beneficial and detrimental. These consequences (both positive and negative) will be unevenly distributed across the participants[2].** We are at the beginning of our experiences with massive data sets, powerful pattern-matching and learning algorithms, and new ways to connect different forms of data. We have some ideas of what might be possible. But, we've spent far less time on what might happen. We know that many of these effects will take some time to become clear. In the interim, some of us will rush to elevate dystopian futures such as "*I am Legend*" or "*Contagion*." Others of us will rush to pursue only the promise. Whatever their view, faculty members, research scientists, doctoral students, and policy makers, will have an eye out for opportunity. We should worry more about subjecting discovery informatics efforts to robust and public (at least open-to-the-scientific community) exploration – such as TREC did for the natural language processing community and why we *began* doing large-scale public trials of medicine.

**Suggestion: Keep discovery local, but link it together better.** My premise is that science is structured to be done in small groups of committed scholars. But, technological opportunities afford larger-scale connections. Let us leverage both. Most science groups are socially-stratified (a few senior people, some junior people, and lots of students). Most group members are co-located (the rise in distributed science is less about discovery than about data management, tool use, and work product-production). Most of these scientific groups have a bounded set of data (see Sawyer, 2008)[3]. While the several large data sets garner great attention, most scientific data is distributed across many computers, protected by many people, and rarely combined. So, discovery informatics tools and approaches that provided for ways to link, to share, and to manage connections with other groups will leverage the technology while enhancing current social arrangements. If discovery informatics tools enabled Napster[4]-like connections this would allow multiple people to benefit from using the tools in ways that were tied to their local practices (e.g., supporting them sharing their data 'playlists' and riffing on each other's work). By leveraging the current social practices and social structures of science in ways that draw in scholars and advances the concepts, approaches and technologies that are currently used to define discovery informatics, you draw people to this opportunity.

## Works cited


Feyerabend, P. (1975) How to Defend Humanity against Science. *Radical Philosophy,* 11: 3-8.

Hughes, T. (1983) *Networks of Power: Electrification in Western Society*. Baltimore, Johns Hopkins Univ. Press.

Knorr-Cetina, K. (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge: Harvard Univ. Press.

Latour, B. and Woolgar, S. (1979) *Laboratory Life: the Social Construction of Scientific Facts*, Sage, Los Angeles, USA.

Sawyer, S. (2008) "Data Wealth, Data Poverty, Science, and Cyberinfrastructure," *Prometheus*, 26(4), 357-371.


[2] This is rediscovered with such regularity and surprise that it may not be mundane.

[3] In saying this, I gloss over the barriers that current data management practices create.

[4] Here I invoke the viral success of peer-to-peer file-sharing system for music that sprang up in 1999.