

I was at SciBarCamp Palo Alto

Martin Fenner, Gobbledygook

July 10, 2009

SciBarCamp Palo Alto took place July 8-9 in the Institute for the Future. I came right from the airport and arrived too late for the general introductions and session suggestions. But there was time for a little break before Sean Mooney started his keynote lecture.

Keynote: Biomedical Research in the Age of Cyberinfrastructure

Sean Mooney

Sean Mooney just recently joined the Buck Institute for Age Research as director of their bioinformatics core. He was talking both about his extensive experience building web-based bioinformatics tools, but also about *Laboratree*, a social networking tool for scientists with a focus on research management. The key argument he was trying to make is that we have a large number of online tools to store and analyze the data of our scientific experiments, but that there is still not enough effort to connect these tools. The infrastructure networks that exist (e.g. caBIG) usually focus on one particular domain (cancer research in this case), and the 27 NIH institutes all have their own approach to informatics. Researchers and administrators have a very different approach to online tools. Whereas researchers want to analyze data, administrators focus more on the much broader picture, e.g. asset management. Administrators often misjudge their needs for the needs of the scientist. Scientists often mistake today's needs for tomorrow's needs.

Sean also talked about his experience building *Laboratree*. He thinks that the social networking for scientists killer app is one that people “have to use” not “like to use”, and that successful tools are simple and start with a preexisting community.

After the keynote we had un-dinner in smaller groups, and I went to bed after being up for 24 hours thanks to the 9 hour time difference. The next day we had the unconference sessions suggested the day before, and I talk a little bit about the sessions I attended. The rest of the sessions are well documented in the SciBarCamp FriendFeed group and on Twitter.

Open Source Textbooks

Chris Patil

This was a great brainstorming session. We agreed that – with very few exceptions – there is no financial incentive for scientists to write textbooks. The main motivation is reputation, and we thought that the lack of return on textbook authoring is an opportunity for new approaches. The concept of Open Access textbooks is a social problem (e.g. who pays for the textbook, what are the author incentives when there is a long list of contributors), and not a technical problem. OpenCourseWare was mentioned as a great educational resource that overlaps in intent with Open Access textbooks. Sean Mooney and Cameron Neylon mention the often unsolved copyright issues when creating Open Access material, and this includes potential issues with proper attribution with the Creative Commons license when reusing material. We briefly talked about document formats (where PDF probably is still important), and that we probably need to print most of the textbooks, at least until eBook readers become cheaper.

Personal Genomics

Melanie Swan

This for me was a great introduction to the topic, and the main emphasis of the session was probably on the consumer perspective. The first ever consumer genomics conference was held in Boston last month. One important, and still only partly solved issue is huge amount of data generated, especially if we take into account not only genetics, but also epigenetic changes, and the fact that not all cells of an individual are genetically identical (which is true especially in cancer). The technology is changing rapidly and the cost of sequencing a whole genome of an individual is constantly coming down. We talked about a few examples of genes that are already useful to test, one prominent example being CYP2C9 and warfarin (an anticoagulant) metabolism. Many chronic disease conditions are multigenetic, and therefore it is usually risk assessment rather than yes/no answers. Some personalized genetics testing companies are Navigenics, Decode, and 23andMe.

Article level metric from the PLoS perspective

Peter Binfield

Peter started the session by addressing the problem: how do you access the worth of impact of journal articles, and at what level do we measure this: journal, research institution, individual researcher? he then went over currently used article level metrics that fall into six areas:

- citation metrics
- usage metrics
- expert rankings (Faculty of 1000)
- Conversations (blogs, media coverage, comments)

- social bookmarking (citeULike etc)
- Other cool stuff (geotagging of authors, etc.)

In this broader context article-level metrics can be seen as post-publication peer review. Much of this is still ongoing, ResearchBlogging.org is for example building an API so that blog posts about PLoS articles can be tracked. In August PLoS One will add information about usage statistics to all their papers. Because usage statistics are currently not widely available, this gives a push to the role of usage statistics in evaluating a paper. Future work at PLoS will include better integration with Mendeley, Zotero, CiteULike and similar services. We briefly talked about the relative little use of commenting and tagging at PLoS One, and some people in the audience felt that we need automated tools (e.g. by looking at what papers are bookmarked or stored in Mendeley) rather than user generated content.

Spinning Science: The good and the bad of media sensationalism

Naomi Most, Kiki Sanford

Naomi and Kiki started the session by putting up the provocative hypothesis that sensationalism is the way to report science. They cite Edwin Slosson who in 1921 said *Our best plan is probably to try to crowd out falsehood by truth and to present scientific information in a way that will be at least as attractive as the misinformation that now holds the field*. They point out that there are different realms of science communication, including

- science for scientists
- science for scientists working in different fields
- science for the educated public
- science for students, and more

We probably need more translators of science, rather than more people doing science. We are already overwhelmed with the massive amount of science created today. We had a longer discussion about the misunderstandings between scientists and their institution's PR people. Part of the problem is that scientists are not automatically good communicators. Should they all become good communicators? This is probably not possible and not required. We then talked about science rock stars and had the best quote of the meeting:

*Why do kids look up more to basketball players rather than scientists?
They are taller. (Jim Hardy)*

Scientific Publishing in 5-10 years

Peter Binfield

This was a fun session of what we think science publishing will look like in the intermediate future. Peter asked the following questions:

- Do current publishers exist?

- Does the journal exist as a package?
- Does the article exist?
- What business models dominate?
- What new technical features do we seriously expect?
- What new modes of scholarly communication may gain wide acceptance?

Other questions that were asked by the audience include:

- What money exists in 5-10 years?
- Who is doing the actual work of publishing?
- What are the customers?

We went through most of these points. We discussed a possible business model similar to iTunes where the per-article charges would be much lower than today. In the future the role of the librarian will change. It will be more about information, and less about journal subscriptions. We talked about how the process of producing and publishing a paper will become cheaper by better tools, and by using the NLM-DTD standard for submission. Articles in the future will not contain only text and figures, and we had a longer discussion on how digital media can be preserved in the long run, and that at some point we will not be able to keep all our data because of the dramatic increase in the carbon footprint required.

Efficiency and incentives in research – How to bend the internet to scientists

Cameron Neylon, Jason Hoyt, Duncan Hull

The last session was really three separate little presentations by the three speakers. Cameron started with a presentation he recently gave at NESTA (Science in Society). One question he asked was how we maximise the efficiency in generating impact for our research? Jason Hoyt gave a good introduction to Mendeley and used this as a basis of what social tools for scientists should look like. One important point: these tools should work with a userbase of 1, rather than needing a critical mass. Duncan Hull talked about Digital Identity on the Web, something he has talked about before. He pointed out that many papers in Google Scholar are by “forgotten password”, “already registered”, etc. OpenID is a solution to some of the problems of author identity, but is complicated to use and not secure enough for some specific scientific applications. Peter Binfield explained that an author identifier would be really helpful for PLoS, and that their databases currently can’t identify an author that has for example changed institutions. We had a long discussion about potential benefits of researcher identifiers, and what is needed to really get the ball rolling.

After the final session I went to the pub with a few people (really strange to do that in Palo Alto), and after having some food hopped on the bus that brought me back to my hotel.