Information Services & Use 28 (2008) 243–246
DOI 10.3233/ISU-2008-0577
IOS Press

ICSTI Public Conference 2007 – Final keynote and close

Final keynote speech

Open access, data-driven science and the impact on research communication

Lee Dirks
Microsoft Corporation

Before going into the future of scholarly communication, let us take a look at the past and see how a hypothetical chemist, Dr. Ramshackle T. Higginbotham, conducted his scientific research around 1857. He has his lab notebooks where he stores his data, and the notebooks are stored on a shelf in his laboratory. To communicate his results, he might publish in a scholarly journal, he certainly belongs to learned societies and attends conferences. Now let us look at his great-great grandson Dr. Ramshackle T. Higginbotham the fourth, perhaps at MIT around 2007. He is also doing research, and puts his data in digital eNotebooks, but his research data is much more complex and when it comes to communicating his results, the channels have really exploded. He is probably blogging and has to read the blogs of his students and colleagues; he might be driving a wiki in his group; he publishes in scholarly journals, deposits his articles in an institutional or central repository in his discipline, attends global conferences and monitors podcasts and RSS feeds as well.

So what are the impacts of this new environment on scholarly communication? First, open access to scientific content and data will become the norm. This does not mean that traditional paid journals will disappear, but the volume of freely accessible data will undoubtedly increase. Second, international cross-discipline research facilitated by interoperable standards will also become the norm. This might be problematic but things such as OAI-PMH, DOIs and OpenURL, which were hard to imagine a few years ago, are slowly becoming the norm and are needed to make sense of the environment we are moving into. Peer review as we know it will be supplemented by technology. Evolved methods of peer review will be adopted, thereby facilitating research quality and authority. Technology alone cannot replace peer review but it can supplement it. Finally, the preservation of data will become a requirement and like Open Access, it will be mandated and handled by funding or government agencies and one can even imagine an ecosystem of private companies providing this service. Lastly, services are going to develop around scientific contents and prevail over the pure publishing concept by providing data analytics, publishing workflow tools, long-term storage and access. *Educause*[1] is an example of the new concepts presented today. It is a program looking at IT in higher education in the United States and it produces an annual "Horizon report" noting the key trends and critical challenges facing higher education over the next five years. One of the issues raised is that academic review and faculty reward are increasingly out of sync with new forms of scholarship. Some way of circling back with scholars and giving them the credit that is due to them is needed.

The new forms of assessment continue to present a challenge for educators and peer reviewers, and, while progress is being made, issues of intellectual property and copyright continue to affect how work

[1] http://www.educause.edu.

is done, and time is lost trying to circumnavigate these issues. It is clear that there is a trend in the direction of Open Access, since 2,621 out of the 23,000 scholarly journals published worldwide are listed in the Directory of Open Access Journals,[2] i.e. a 10% growth in a short period of time, and the trend is continuing. Some declarations and mandates are pushing in this direction in the United States and Europe. Talking about the impact of OA, a PLoS article says that: "publicly available data was significantly associated with a 69% increase in citations ..." and "the correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data". However, the increased load on peer review might turn into a quality issue and this is where technology could supplement the peer review process as we know it. The Web 2.0, which some see either as positive or negative, might be a double edged-sword. Web 2.0 technology could have an impact on peer review with numerous functionalities baked into journals and the scholarly communication process whereby people can provide their feedback in real time therefore opening up the peer review process. Blogs are a fascinating twist because they can record not only successful experiments but also those that failed and open real-time discussions in the blogosphere on the reasons of the failure. *OpenWetWare*[3] gives small labs the opportunity to able to share best practices in real time, and *Nature's Connotea*[4] makes it possible to share research, see what others are looking at and share it in a very collegiate way.

Nowadays, there are over 1,400 repositories in the world while 10 years ago there were almost none. By taking advantage of the different types of repository software available to capture information, not only university campuses but also corporations and government agencies have been developing repositories. However, the availability of such software in an institution does not necessarily mean that it is used. So repositories are not problem-free and there are issues as to what should be deposited and how it should be used. *Webometrics*[5] has been keeping track of what is being stored and used in academic institutions and ranks them according to Google ranking. It clearly shows the impact on ranking that publishing in an institutional repository has. *Outsell*,[6] an industry monitoring company in the United States, recently published a report on institutional repositories showing that, although depositing is not has high as expected, when that behavior changes and self-archiving becomes a common practice, it will have implications for OA and traditional publishers. The concept of advertising might be distasteful to some, but a market research by *Outsell* on users' acceptance of advertising in OA journals did show that "a full 80% indicated some degree of willingness to consider ads directly tied to scholarly articles, with 49% indicating direct willingness". These results are quite surprising but they certainly open up possibilities.

Data sharing is the single most dramatic change that will happen in scholarly communication in the coming years and it will radically change science. The benefits of data sharing are obvious but there are some important issues such as integration and interoperability between silos of information, data from dataset to atomic levels, and data provenance and quality. Another issue is methodology and the ability to export and use accepted standardized formats. Lastly, there is the security issue and the ability to lock down or provide read/write access to all or parts of the data. The ability to provide advanced data sharing functionalities such as "live documents" do with links to simulations and raw data, will make a difference in the market place. Microsoft is currently looking into live documents with some East Coast institutions. Even corporations such as Novartis are starting to realize the importance of data

[2] More than 3,600 journals in September 2008.
[3] http://openwetware.org.
[4] http://www.connotea.org.
[5] http://www.webometrics.info.
[6] http://www.outsellinc.com.

sharing in the interest of the community. Another issue that remains is the way institutions and nations can improve this sharing trend. The NSF in the United States and the DRIVER project in Europe are working on a cyberinfrastructure but some problems such as interoperability have yet to be solved. In the United States, some agencies are working together to store all scientific data generated by federal agencies in publicly accessible repositories.

The ultimate in peer review is the ability to take your colleagues data for a test spin. Many tools for data analysis are being developed such as *Swivel*,[7] *Many Eyes*,[8] *Gapminder*,[9] and *Freebase*[10] and in another vein CSA's *Illustrata*.[11] Those are the services that will help develop the new scholarly communication ecosystem and advance science. *Many Eyes* is a data visualization tool available from IBM and *Swivel* is produced by an independent company. These tools are currently charge-free, but there is no doubt that down the line some levels of services will be subject to charges. Once again this raises the issue of quality and often key science sites fail to register in the top 30 Google search results and this is an important concern for all involved, from institutions and government agencies down to citizens. Standards accepted by disciplines and institutions will be needed to control the proper use of datasets, and metatagging methods will be crucial down to the atomic level. The publishing ecosystem will shift with the ability to add value with services and many companies will move into the ecosystem to provide solutions based on open software. With the Web 2.0 and this ecosystem shift, prototypes will be rapidly available in alpha versions to gather feedback from the community.

Before concluding, I will show a few slides of Neptune,[12] a joint project of Microsoft, the University of Washington and several Canadian universities, to illustrate e-Science mash-ups. This project consists in putting a sensor network off the west coast of the United States and Canada to monitor not only seismic events but also the impact of undersea volcanic activity on wild life. This network will be connected and controllable over the Internet. It will be possible to pull data from the oceanic floor, compare it to data from various institutions including NOAA (National Oceanic and Atmospheric Administration), run it as simulations across grid networks, map it out and, once analyzed, disseminated it to peers.

To conclude, one might ask who is going to be in charge of caring for the data to be fed into data infrastructures. It could be publishers or third parties but wouldn't it be a great opportunity for libraries to step into the breach and take advantage of technology to move the traditional concepts of data storage, preservation, management and dissemination into the new scientific information ecosystem. When it comes to where scholarly communication will be in five to ten years, it is clear that open access to both text and data will be the rule and not the exception. Publications will be live documents with links to real-time data and related software. New forms of peer review including social networking will have been accepted and adopted. Blogs and wikis for collaborative research will be normal operating procedure, in many cases they already are but even on a global scale this will become a predominant approach. National and international repositories will be a key part of the scientific cyberinfrastructure and the US and Europe are already moving in that direction. Preservation and access to datasets will be a mandated part of the scientific lifecycle and a service industry will develop around online data analysis, visualization and dissemination of scientific information.

This is exciting but there is still a lot of work to do.

[7] http://www.swivel.com.
[8] http://www.many-eyes.com.
[9] http://:www.gapminder.org.
[10] http://www.freebase.com.
[11] http://www.proquest.com/promos/product/csaillustrata.shtml.
[12] http://www.neptune.washington.edu.

Summary and close

Raymond Duval
INIST Director

Ladies and Gentlemen, dear Colleagues,

First, I would like to tell you that Arnold Migus, the director of CNRS, had planned to be here today for the closing remarks. Unfortunately, he had to cancel his trip to Nancy and he asked me to express his apologies. However, Mr. Migus is very interested in the issues that were discussed during this conference and the CNRS will be fully involved in addressing such issues. For me, it has been a great pleasure to welcome the ICSTI Public Conference here in Nancy. The topic of the conference was especially timely. Today, research assessment plays an important role in the relationships between the definition of national scientific policies and the challenge of increased globalization.

As players in the scientific information field, all of us are concerned by research assessment because as researchers we are authors and information producers, as research organizations and funding agencies we need to assess the impact of scientific policies, and as publishers and information professionals, our goal is to facilitate changes in assessment methods and tools. We are not here to question the current system. It might have limits related to the over utilization of bibliometrics as some speakers have pointed out, but it gives institutions useful quantitative data. We are here to get the most out of the best international practices so that we can adapt out current methods to changes imposed by a knowledge sharing society. This is why I feel this event can promote ideas and help reach this goal. Because of our position, we have a good opportunity to observe the impact of digital technology on research. In recent years, we have seen far-reaching changes in scientific communication and we have learned a lot from these changes. First, we have learned that assessment methods that apply in some disciplines are not relevant in others because the communities have different practices. Second, we have learned a great deal because technology has somewhat matured. Today, rather than relying on citations alone, we can assess the impact of research on a given community by gaining insight into usage. And third, we are also aware that Open Access offers new assessment possibilities.

To conclude, I would like to thank the speakers and participants and especially those of you who traveled a long way to come to Nancy. I also want to thank all of you for the variety of approaches expressed during this conference and the quality of the discussions we heard. I also hope that you appreciated our welcome as much as we appreciated your presence. And I would like to thank Herbert Gruttemeier for the organization, and also the ICSTI secretariat and the INIST Communication Unit.