

Open Access, Open Data: paradigm shifts in the changing scholarly communication scenario

Elena Giglia, University of Turin

The expert conference [Open Access Open Data](#), held in Cologne (D) on December 13th and 14th 2010, represented a two-day intensive debate and exchange occasion. Excellent [contributions](#) by the invited speakers alternated with sharp and stimulating participants' interventions, in order to create an inspiring atmosphere. The common scenario being the changing scholarly communication system, we can try to follow some key insights emerging from the debate.

Scholarly communication, digital age and the commons

Dieter Stein (Düsseldorf University, D) highlighted the free nature of scholarly communication, and the yet unexploited potentiality of the Net: the concepts of access, visibility, impact, should have to be revisited in the perspective of an Open Science, where the "User Generated Content" calls for more "liquid" and "current" publications. Open Science stands for transparency and efficiency, made possible by the new ways of producing science, in which one can discuss and debate in the meanwhile science is processed, and not only at the end of the work, in front of a final "product". This new attitude should claim also for a revisited concept of copyright, because in the digital age there are no more paper objects to be possessed by anyone.

On this path, **Rainer Kuhlen** (Konstanz University, D) developed a fascinating theory moving from the concept of knowledge as commons and coming to a "knowledge ecology", where sustainability of immaterial goods can only be achieved by open and free access and unrestricted use – on the opposite, natural resources are to be made scarce to be preserved. The knowledge ecology is aimed at achieving the goal of people-centered, inclusive and sustainable knowledge societies. In the "Commons" paradigm, this "knowledge ecology" - and in its context the idea of open access - provides an alternative both to existing commercial publishing models on the international information markets and to international copyright regulations: they both have mainly emphasized only the economic impact of knowledge and information. But it should also have to be taken into account the genuine character of knowledge as a common-pool resource. In this perspective, the "knowledge ecology" does not object to the commercial use of knowledge produced in public environments such as universities and research centers, but suggests that publishing models are only acceptable when they acknowledge the status of knowledge as a commons, allowing free and open access for everyone. This commons must be based on sharing knowledge, producing new knowledge collaboratively, and providing future generations with the same access and usage rights. As a result, the commercial use of publicly produced knowledge should be the exception and open and free access the default; there should be a need for compensating the public when the commons "knowledge" is exploited for commercial usage; there should be a need for new property right rules when knowledge is increasingly produced collaboratively: nobody should have an exclusive right on knowledge, neither the author, nor the publisher, nor the reader; there should be the need of new sustainable business models compliant with such a scenario. Access is the real challenge, as knowledge is usable only if accessible.

Alma Swan (Key Perspectives and Enabling Open Scholarship, UK) stressed this concept of "access" and declined it according to her concise, factual and strong definition of Open Access as «immediate, free (to use), free (of restrictions) access to the peer reviewed literature and data». "Immediate" means that no embargo is suitable with Open Access. On the other hand, advantages in Open Access are so many that

both the single researcher and the institutions have to gain from the increased visibility and impact deriving from this new paradigm. Many and many examples in different disciplinary fields were showed, alongside with evidences aimed at removing the common misunderstandings about Open Access itself. Open Access increases citations and boosts the flow of ideas by immediately disseminating the papers, and it gives visibility to works before closed in not-circulating paper journals: download statistics of the Institutional Repository ORBi, supported by a strong Rectoral mandate at the University of Liege (B), show that log files come from many Francophone countries which never had accessed before the same journal articles. Potentially, by connecting readers worldwide, Open Access fosters international and interdisciplinary collaborations; it ensures access to academic findings also to Small and Medium Enterprises, where innovation is carried out. Open Access is actually different from vanity publishing or the “stick anything up on the Web” approach. Open Access means «moving scholarly communication into the Web age», which is in perfect consonance with all the voices heard during the conference.

Open Access and Open Data

Who better than **Stevan Harnad** (Southampton University, UK – Montreal University, CAN), one of the worldwide most known Open Access advocates, should have discussed convergences and divergences between the two topics referred to in the conference title? Harnad strongly supported the idea that self-archiving (“Green road”) is immediately feasible, results to be the most cost-effective strategy, and relies only upon the authors’ will, which can (and ought to) be supported by institutions’ or funders’ mandates – the latter enforcing themselves by the positive metrics in terms of downloads and citations. The free, immediate and permanent full text online access stated by Open Access definitions refers to peer reviewed literature. Access to any other academic material, data included, is desirable but it’s not the focus of the Open Access paradigm, moreover, as it’s not immediately feasible, it risks to delay the whole process. Datasets are similar to journal articles, as they often speak for themselves, but whilst the final work gains from being disseminated, often researchers should have been granted the right of first exploitation of their data. That’s why Open Access can be mandated, and Open Data can’t. Besides, science and scholarship is quite different from data gathering – it is supposed to be data interpretation. Once mandated Open Access self-archiving, as “Green” Open Access grows, data-archiving too will grow, because of their natural complementarities and the power of global collaboration to accelerate and enhance research progress. Data too, as journal articles drafts, should be deposited locally in Institutional Repositories and then harvested centrally by subject-based repositories, which seems to be the most cost-effective solution in order to maximize the disseminations of research’s outputs.

Malcolm Read (JISC, UK) agreed on the distinction between Open Access to research’s outputs (that can be mandated) and Open Data. JISC requires data to be available and correctly stored for ten years since their publications, but not made openly available, at least till now. Data are a competitive advantage for the researchers who gathered them, so a right of first exploitation and mining should be provided. “Open” in the JISC perspective is a key issue, a framework in which strategic topics like Open Access, Open Data but also Open Educational Resources can be addressed upon the principle that publicly funded research ought to be publicly available. In the UK has been created an Open Access Implementation Group, chaired by prof. Martin Hall (Salford University), to coordinate the efforts of all the involved actors (academic institutions, funders, Research Councils). Next step towards Open Data should take into account keywords like integrity, re-usability, long term preservation, trust, sustainability. Guidelines implementation is needed as well as the creation of a seamless infrastructure for storage, search and retrieval which can rely on the existing net of Institutional Repositories. Along this path, some of the major concerns expressed by researchers are their lack of expertise in treating, metadating and preserving data, the presence of legal

and ethical constraints, and the fear of inappropriate use of the datasets. Yet, the effort is worth, because of the recognized significant potential in data sharing with other countries like USA, Australia and Asia under the umbrella of the forthcoming European 8th Framework Program.

About the ongoing 7th Framework Program and its Open Access Pilot Project talked **Celina Ramjoué** (European Commission, DG Research and Innovation). She gave an overview of the roles of the European Commission as a policy making body, as a research funding body and as a capacity-building body, which turn in a global approach aimed at improving access and dissemination to foster progress, at enabling innovation through improved access, and at increasing returns on investments on R&D. The European Research Area (ERA), as stated in the Lisbon Treaty, art. 179, envisages the so called “Fifth Freedom”, i.e. the freedom of circulation for researchers, scientific knowledge and technology. Open Access is a strategic keystone in delivering ERA, as reckons the [Digital Agenda for Europe](#), COM (2010) 245. Open Access to publicly funded research is a principle stated also in the Innovation Union, COM (2010) 546, both in Commitment 4 and 20, in order to promoting openness and capitalizing on Europe’s creative potential. In this perspective, the EC is actively engaging all the member states with questionnaires and the recent Workshop held in Brussels on November 25th-26th to find out the next steps towards a sound strategy to Open Access: among the others, the need for mandates, the need for interoperability and standards, the need for a specific copyright law for scientific publications, and the issue of Open Data. The recent Report of the high-level expert group, [Riding the wave: how Europe can gain from the rising tide of scientific data](#) (October 2010) is a further step toward a 2030 vision in which « All of these principles – our vision - point in the direction of an infrastructure that supports seamless access, use, reuse, and trust of data. It suggests a future in which the data infrastructure becomes invisible, and the data themselves have become infrastructure – a valuable asset, on which science, technology, the economy and society can advance.»

One of the keyword of the debate – reuse – was highlighted by **Sünje Dallmeier-Tiessen** (CERN, CH), both in the perspective of research integrity and in the vision of an acceleration in knowledge creation generated by data sharing. Different scientific communities have different prerequisites and specification in dealing with data, that’s why tailored models are supposed to meet the different needs. The contribution of information specialists and their expertise is fundamental in achieving a proper data treatment.

Publishing, searching and finding data: what’s new?

During the two-day conference, some key issues in data treating have arisen and many current and innovative practices have been shared:

- **New balances between record keeping and knowledge transfer:** in the print age, both of these tasks were assigned to publishing. In the Web age, with a new article issued in PubMed each 36 seconds, the exigencies of record keeping and the credit/acknowledge economy which drive the existing publishing market could no longer meet the exigencies of knowledge transfer, simply because there is too much to read. **Jan Velterop** (Concept Alliance) proposed a fascinating model of “nanopublications”, based on triplets of concepts describing the content. They are at the same time assertions and references which provide an immediate label for a paper but also allow a serendipitous discovery of unexpected correlations between different and apparently non-related works.
- **New integration between data and scholarly communication:** presenting the current landscape of scholarly communication in Germany, **Anita Eppelin** (German National Library of Medicine, D) underlined two different methods in order to design a future scenario: the institutional and the

“grassroots” approaches. As to the first, a Commission for the Future of the Information Infrastructure was set in 2009 in Germany, working on licenses, Open Access, Open Data, digitization and long-term preservation, virtual research environments. In the second case, we have scientists-driven, bottom-up, promising projects such as [Pangea](#) - Publishing Network for Geoscientific & Environmental Data, the [European Virtual Observatory](#) and the [European Southern Observatory](#) in Astronomy, [ArtsHumanities.net](#) as a hub for digital Humanities, [CESSDA](#) - Council of European Social Science Data Archives, an umbrella organization for social science data archives across Europe, and the new challenge of data publishing. The underlying hope is that Open Access and Open Data might promote each other.

- **Data publishing:** although scholarly communication is in everyone’s opinion more and more data driven and data intensive, as to now data are not always considered as in integral part of a scientific article, as both **Olaf Siegert** (German National Library of Economics, D) and **Gert G. Wagner** (German Data Forum, D) highlighted. They are often linked in an external database without any connection to the original article. **Martin Rasmussen** (Copernicus publ.) presented both some examples of scientific journals linking the article and the dataset via DOI and a new innovative Open Access publication such as Earth System Science Data (ESSD), born in 2009. Its scope is to publish articles on original research datasets, furthering the reuse of high quality data. A competitive advantage of this journal is the adoption of a specific peer review on data. As all the Copernicus publications, it applies the Public Peer Review process: the article is accepted after a rapid preliminary review, put in special section on the journal’s homepage – being at that point already readable and citable –, then submitted for 8 weeks to the readers’ comments (“public peer review”) and to traditional reviewers, and eventually revised by the author on the basis of the received comments and published into the journal issue. This process is a solution saving both the needs of a rapid publication and a thorough examination, combining the traditional peer review with the immediate dissemination of the Open Access paradigm.
- **Findability:** how many datasets are already available on the Web but are somehow lost in cyberspace? **Toby Green** (OECD) gave a useful example of integration of books, journals, but even datasets and tables in a single search engine within the OECD library, with an equal bibliographic status. **Jan Brase** (German National Library of Science and Technology) claimed for libraries to open their catalogues to non-textual materials, and presented [DataCite](#), a consortium of 12 institutions worldwide aimed at promoting the use of persistent identifiers for datasets. By assigning DOI names to data sets, data becomes citable and can easily be linked to from scientific publications: data integration within the text is an important aspect of scientific collaboration.