

## OPEN ARCHIVES AND THEIR SIGNIFICANCE IN THE COMMUNICATION OF SCIENCE (WORKSHOP IN UPPSALA 16-17 NOVEMBER)

Urban Ericsson, Jenny Ericsson, Linda Åström Wennbom

### Introduction

Today there are several projects concerned with how to use usage data from open archives in order to understand the new ways of how science is communicated. The key issue in this scientific field is to develop secure and reliable statistical analysis for the usage of the objects deposited in open archives. With the increasing efforts to evaluate science in different ways it is important to understand the metrics from open archives.

At the Swedish University of Agricultural Sciences (SLU) a two-day workshop was held on the topic in November. The aim was to examine how usage metrics can provide new knowledge about how science is communicated through open archives and if it also can be a tool in the future evaluation of science. The workshop was financed by the OpenAccess.se program and we were more than 60 participants taking part of interesting lectures and discussions. This was a multisited workshop. Most of the participants were in Uppsala but some of us participated via video from SLU-sites in Alnarp and Umeå. Some of the presenters also connected via the same video link during the workshop.

The workshop started with a welcome address by *Jan Hagerlid* (National Library of Sweden. Coordinator of the OpenAccess.se programme). In the *Welcome address: Open Access developments in Sweden – why usage metrics matter to us* Hagerlid stated that Sweden has a well-developed infrastructure of Open Access repositories and there is a high official consensus on the importance of Open Access, manifested in OA mandates from several research funders as well as universities. The National Library of Sweden is promoting and coordinating the work for Open Access (OA) in close cooperation with other stakeholders. An obstacle to the OA development is the increasing pressure on researchers to adapt their publishing behaviour to traditional bibliometric indicators. This situation makes the need for standardised and reliable usage data from open sources all the more urgent.

### Speed – Visibility - Usage – Impact

In a very interesting lecture from *Anne Gentil-Beccot*, librarian at the CERN Scientific Information Service,

we learned that in the field of High Energy Physics (HEP) the tradition of sharing science is not a new concept. Already in the 50s and 60s researchers exchanged and communicated their results (before and after peer review) by mail because of the need of a more rapid exchange of thoughts than possible through the traditional publishing via journals. This is the background to the famous arXiv repository founded by Paul Ginsparg in 1991 and the pre-print culture is still strong in the field of High Energy Physics and journals are no longer the main communication tool in the field of HEP (although all papers are still submitted to journals, because journals still have a crucial role: ensuring the peer review). INSPIRE, which is an infrastructure containing the whole HEP literature will bring the scholarly communication to a new level Gentil-Beccot said, bringing in more metrics (from OA and traditional published sources) to the HEP community.

In several of the following speeches we could see that OA publishing were beneficial when it comes to citations compared to non-OA publishing and that the citation gap was increasing over time.

*Stevan Harnad* (Canada Research Chair in Cognitive Sciences, Université du Québec à Montréal, Canada, and School of Electronics and Computer Science, University of Southampton, UK). His speech *Scholarly/Scientific Impact Metrics in the Open Access Era* gave insights into the bigger picture as well as interesting details about metrics and OA. In the online era, potential metrics have extended from publication counts, journal impact factors and citation counts, to include download counts, growth and decay rates for metrics, co-citation measures, etc. Still missing today, however, are among other things: a validation of the metrics, discipline by discipline, that tests and confirms their meaning and predictive power, especially in research assessment, and a sufficiently large and open web wide database to allow the global research community to test, validate and monitor its metrics (which are currently collected systematically only by proprietary commercial databases).

The OA movement is helping to generate the requisite OA database for articles. OA not only makes it possible to harvest research impact metrics web wide, but it has also been shown to increase them (the "OA Impact Advantage"). Harnad also mentioned a new

study (Gargouri et al, 2010) that showed that mandated OA provides just as much of an advantage as self-selected OA. This disproves the self-selection theory, which claims that OA materials are cited more often because authors choose to make their best work available through OA.

The infrastructure is in place – most institutes of higher education now have their own OA repository – but the content is still largely missing. Between 5 – 25 % of research articles are actually being deposited. For OA metrics to improve, we need more content. To get more content we need more OA mandates from funders and universities.

There is now plenty of evidence to show that disseminating research outputs through open archives has advantages in many arenas. In *Alma Swan's* (Consultant working in the field of scholarly communication. Director of Key Perspectives Ltd.) speech, *Visibility, usage, impact, economic benefits – the significance of open archives for research and elsewhere* we learned that research moves faster and more efficiently with Open Access. Open Access means that scientific results gain greater visibility and impact and lead to better possibilities for monitoring and assessment of science, and opens up for new semantic techniques (data and text mining). For authors, advantages include visibility, usage, impact, personal profiling and marketing, as well as research advantages. Alma showed us several testimonies from scientists who had noted new audiences for their research. For institutions the advantages are much the same, on a larger scale - a big OA repository leads to a big web presence. Swan used the University of Southampton as an example. In rankings like *University Metrics - Global University Rankings*, based on the G-factor (counting other universities' linkings to the own universities website) and *Webometrics* the University of Southampton is highly ranked. It is reasonable to think that the early OA-mandate at the university and the now large volume of research output in the repository is a key factor in this ranking, Swan said. In one graph Swan showed how the Queensland University of Technology (QUT) since their OA-mandate in 2004 has outperformed other Australian universities when it comes to research income. Swan said that we cannot be certain of the correlation of research income and OA-mandate. But the fact that much of the increased research income comes from industries and commercial sources suggests that these stakeholders could be finding the QUT research output in the QUT ePrints archive, and author testimony has supported this. Swan concluded that making research openly available increases its visibility, naturally, leading to greater usage. Enhanced usage can then result in enhanced impact for the work, measured in conventional terms by citations from the research community, and also in other ways in society outside the research community itself such as relations

between universities and industry.

With Swan's examples from the QUT ePrints archive in mind it was interesting to listen to *Tom Cochrane* (Deputy Vice-Chancellor, Technology, Information and Learning Support, Queensland University of Technology). His presentation on *Open Access - the advantages for a university of a successful policy - some evidence* showed the contribution that an open access policy and a practice of open access makes to the improvement of the profile and impact of research, based on both quantitative and qualitative evidence at the Queensland University of Technology (QUT). QUT has had an institution-wide OA policy since 2004, mandating deposits of post-peer review articles and conference manuscripts into the QUT ePrints archive. According to Cochrane, a successful policy is only possible if it is attuned to researcher motivations. The implementation of the policy by the university library has focused on feedback to researchers, for instance by providing download statistics and personal pages for each researcher. Another success factor is the high-ranking search results of archive posts in Google which brings greater visibility for the author and the institution. There is growing evidence of research advantages for researchers and the institution as a whole. Cochrane presented several graphs on an individual author-level that showed a surge in citation rates after an author had deposited their articles in the QUT institutional repository. There is also qualitative impact evidence coming from the researchers themselves. Cochrane presented statements from individual researchers who had reached new audiences in the developing world, noticed a greater exposure to industry and also found that their own students took an interest in their teachers' ePrints documents.

In several projects added values are implemented to the research objects because of the data produced when a user visits and uses information from the scientific object. *Sergey Parinov* (Central Economic and Mathematics Institute of the Russian Academy of Science, Russia) gave a lecture on the Socionet project, a Russian research environment connected to open archives. Under the heading *Advanced communication of Open Science: the Socionet approach* Parinov talked about the importance to re-use scientific objects and the concept of liquid publications, meaning that a research object is not a fixed entity. Fragments of articles are re-used, circulated, cited and built upon and these aspects are something the implementation of the Socionet infrastructure visualizes. Within the Socionet basic research assessment statistics is provided by automated monitoring services which trace all relevant changes in Socionet scientific data, Parinov said and added that the service collects data to keep updated statistical portraits of all registered research results/outputs, researchers and organizations.

Another interesting national project was presented by *Thomas Severiens* (Department for Mathematics and Computer Science, University of Osnabrueck,

Germany). In the talk *OA-Network: An infrastructure layer for enhanced visibility services* Severiens outlined the German project Open Access Network, funded by the German Science Foundation (DFG) and coordinated by DINI the German Initiative for Networked Information. These projects are now building up an infrastructure layer to enhance the visibility of OA research publications. The presentation gave an overview about these projects when it comes to collecting, enriching and distributing metadata and data from repositories. Severiens visualized how different services relied on open archives and their metrics in order to bundle objects/usage together for the benefit of the author.

*Paul Needham* (Cranfield University, PIRUS2 Project Manager) and *Peter Shepherd* (Project Director of COUNTER) talked about *PIRUS2: Developing Practical Standards for Recording and Reporting Online Usage at the Individual Article Level*. In the PIRUS2 project the aim is to address technical, organizational, economic and political issues related to usage statistics on an individual article level. By doing so it also aims to specify standards, protocols, an infrastructure and an economic model for the recording, reporting and consolidation of online usage of documents hosted by repositories, publishers and other stakeholders. Gathering, consolidating and re-exposing are aspects of interests to the project and the metrics from open archives are crucial for the understanding of how science is communicated. This service needs a business model. The project will present some possible models early next year said Needham.

## Summary

There are promising futures for the open archives in the field of scientific communication. All speakers brought up the benefits of open access when it comes to aspects of Speed – Visibility - Usage and Impact. The full access to scientific publications available from repositories is now fundamental for the evolving services we see in different innovative projects.

## Reference:

Gargouri Y, Hajjem C, Larivière V, Gingras Y, Carr L, et al. 2010 *Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research*. PLoS ONE 5(10)  
<http://www.plosone.org/article/citationList.action?articleURI=info%3Adoi%2F10.1371%2Fjournal.pone.0013636>

With the opportunities to track the re-use of scientific objects from open archives we are heading towards new interesting metrics. These metrics can tell us a lot about how an object is used and give feedback both to the author, the institution and the re-user. Some of these new services were described as bundling, i.e. aggregating information from different repositories and objects.

We also heard of the scientific object as being liquid, meaning that bits and parts of the object moves around still with the ties to the original object but being re-used and tracked when it shows up in new contexts. Exciting and promising futures were outlined during the two-day workshop which will give us new understandings of the field of scientific communication.

As services are evolving and infrastructures are on the way you get the feeling that the future of the scientific communication landscape is already here. But there is one crucial thing missing and that is content, as Harnad pointed out: The major obstacle to OA development is the fact that OA's target content – refereed research articles – is not being deposited; without it, usage metrics are of limited usefulness. What is needed is deposit mandates. Once there is OA content, rich metrics will evolve quickly and naturally. What today's sparse metrics can be used for is only as an incentive, showing the benefits of OA deposit to depositors and prospective depositors.

The speakers at this workshop all made it evident that research and researchers gain from Open Access. OA published articles are being cited more than non-OA and that gap increases over time. Therefore, we all must turn up the tempo in terms of universities' and funders' OA mandates