BOM@ERCIM
Towards an open access policy for ERCIM

J. Baeten[1], L. Candela[2], I. Fava[3], C. Kirchner[4], W. Mettrop[1], L. Romary[4], L. Schultze[1]

[1]Centrum Wiskunde & Informatica, Amsterdam, Netherland

[2]Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", National Research Council of Italy, Pisa, Italy

[3]National Research Council of Italy, Rome, Italy

[4]Inria, France

Abstract

This working paper presents collects background elements that have been discussed when designing the set of recommendations expressed in the BOM report. Such background elements highlight the key role current and future repositories may play in the open access arena.

Contents

Introduction and background

The wide dissemination of research results is an essential component of a prosperous development of science and technology. The legacy of the scientific publishing landscape, created in the print era, as well as the prospect that the wide deployment of information technologies may lead to a simplified dissemination of research assets, have made it necessary for researchers, libraries and all institutions involved in higher education and research to contribute to an **in-depth evolution of the scientific communication processes**.

Without drawing a complete history of the open access movement here, we can identify some major milestones, which have lead to the current situation:

- Dramatic increase in journal subscription costs in the 90's (so called "serial crisis" (Panitch and Michalak 2005));Setting of the first publication repositories with arXiv in 1994 (McKiernan 2000), following several years of pioneering dissemination of digital papers on ftp sites, mailing lists etc.;
- Emergence of the Open Access movement stating that scientific content, seen as a public good, should be freely accessible and re-used. We can mention here the Bethesda Statement on Open Access Publishing[1] and the Berlin declaration issued in 2003 and signed by many scientific organisations worldwide;[2]
- Further deployment of institutional, thematic or national repositories worldwide (see Romary & Armbruster, 2006) contributing to the expansion of so-called *green* open access;[3]
- Rise in the number of native open access journals (*gold* open access)[4], accompanied by the emergence of new publishing models based on the payment author fees (or APC, Article Processing Charge), where scholarly papers are made available online at the time of publication.

Behind the notion of public good, we can also mention the importance of **scientific, institutional or national sovereignty**, whereby publicly funded research should lead to productions that remain within the remits of their producers. This has direct consequences at various levels: preservation of authors' copyright, setting up reliable infrastructures, defining publication costs in accordance to clear services provided by third parties (as opposed to the current non-competitive publication market), with the objective of increasing the capacity of public actors to carry out their assessment or strategic planning duties in full independence.[5]

In Europe, green open access has grown substantially, as is shown in OpenAIRE (www.openaire.eu) that collects the content of many individual European repositories. And France, for instance, has developed a national publication repository platform: HAL.[6] At the national level, the French Minister for Research issued in January 2013 a strong statement on open access to scholarly results, where the archiving of publications in open access repositories, and in particular the HAL national platform, has been given priority, in

---

[1] http://legacy.earlham.edu/~peters/fos/bethesda.htm.

[2] http://openaccess.mpg.de/Berlin-Declaration.

[3] Green open access can be defined as the free dissemination of scholarly papers by *authors* themselves.

[4] The Directory of Open Access Journals DOAJ lists more than 10,650 Journals from 135 Countries in October 2015.

[5] For instance without relying on the sole Journal Impact Factor.

[6] https://hal.archives-ouvertes.fr.

complement to the definition of new viable editorial and business model for scientific publishing.

Some institutions such as Inria or CWI have managed to go even further by **mandating researchers to deposit publications within their repositories**. Such mandates are usually linked to annual assessment, whereby the papers in the repositories are taken as the sole source for reporting by researchers or research teams.

Initiatives are also taking place to develop new publishing models that decouple the notion of publication, understood as the dissemination of the content online, from the peer-reviewing process. We can mention here the Episciences initiative in France (Berthaud et al. 2014) that offers an environment for implementing so-called **overlay journals** on top of existing publication repositories.

The ERCIM partners represented in the BOM@ERCIM group have expressed the need for a better communication between research performing organisations and at the ERCIM level, the provision of a strong set of recommendations that could step by step be implemented among all ERCIM members. The present document provides the basis for such a policy.

Providing access to our publications through institutional repositories

Publications are primary research outputs, especially in the context of scholarly communication. However, movements like "Open Science", "Science 2.0" or "data-intensive science" are pushing for the "publishing" (in the sense of "making public") and "sharing" of the entire spectrum of research "products"/"records" that are produced during a research activity including papers, datasets, presentations, workflows, software. This expectation is leading to a number of changes in the scholarly communication arena, e.g.: new "publication typologies" are envisaged (Bardi and Manghi 2014), data journals are proliferating (Candela et al. 2015), repositories (Romary and Armbruster 2010) are offered "as-a-Service" to the entire scientific community.[7]

In this dynamic and evolving environment the need to adapt and share practices, approaches and tools promoting an "open" access to research products across the boundaries of scientific domains and organisations remains a challenging topic. It is so challenging that, even when only considering the well-known publication domain, it calls for precision and development.

Making assessment easier for the researchers and their institutions

Research assessment is part of the core institutional activity of a research organisation. The experience of many universities and research institutions (e.g. the University of Liege)[8], who are currently using the Institutional Repository as the basis from which to gather information to assess research publications, appears to be a good practice to take as an example. ERCIM members should develop **institutional policies binding research assessment to publications that are deposited in the Institutional Repositories**, establishing some forms of rewarding authors whose publications are available open access (either deposited in the institutional repositories or published in OA journals).

---

[7] OpenDOAR (http://www.opendoar.org), re3data.org (http://www.re3data.org), BioSharing (https://biosharing.org) are just three examples of Directories of existing repositories.
[8] https://www.ulg.ac.be/cms/c_17700/en/open-access.

Recommendation: ERCIM member organizations should detail in their policies that assessment activities (reporting) are closely linked to the presence of papers in their Institutional Repositories; papers should be available in Open Access form when possible.

Immediate access to publications

Researchers are best served when their results are disseminated immediately and a majority of funders also demand immediate open access to all research materials resulting from this support. The open access movement promotes unfettered access to all research materials. Conventional publishers have only invented (long) embargo periods to minimize the effect this green OA has on their sales. At the moment we are clearly in a transition period from a pay-wall system created by conventional publishers to some form of global open access in the future. Options are immediate green, gold with the involvement of traditional publishers or a completely new way of gold without traditional publishers (cf. overlay journals, PLOS one). Either way, **embargos are not relevant** in the context of this transition and the aim should be to have all research materials available in immediate open access.

All options for the future have pros and cons.

Green: all OA, possibility to disseminate post referee version in a repository, but in some cases embargoes from publishers will have to be taken into account and pay walls will continue to exist.

Gold OA journals with conventional publishers: no more embargoes and the peer review process remains with the publishers. Costs for APCs replace license costs and will still be dictated by the publishers.

Gold OA journals without publishers: no embargoes, an infrastructure should be built to handle the review process. The research community takes publishing into its own hands. Initial costs to realize the infrastructure and review platforms but costs will be minimal and more controlled in the end.

Recommendation: During this transition period all ERCIM institutes should be committed to disseminating its research as soon and as widely as possible. Copyrights ideally are in the hands of the institutions (or authors) and a digital copy of the author's version should by default be placed in an open access repository no later than the date of publication. Embargoes from publishers should not be accepted.

Simple licences to facilitate re-use

Provided that authors keep their authors rights on their research output, proper licenses should apply in order to facilitate the reuse of the above-mentioned resources. A recent study by OpenAIRE project (Guibault and Wiebe 2013) examined licenses to be used both for licensing publications and data:

"In principle, all six possible combinations of the CCPL[9] licence meet the open access principles of free accessibility, further distribution, and proper archiving. All CCPL combinations also guarantee proper attribution of the author, in application of the mandatory "Attribution" license element; the No-Derivative license element allows the author to permit only verbatim redistributions of the work; the Share Alike element allows

---

[9] Creative Common Public Licence (https://creativecommons.org)

the creation of derivatives as long as the same or an equivalent license is applied to the final work. The upcoming version 4.0[10] of the CC-licenses will include in the scope of the licence not only copyright, but the sui generis database protection right too. This makes this license finally suitable to license not only scientific publications but also the related research data, all with one single instrument".

As Creative Commons Licenses are widespread and well known by researchers and publishers, the suggestion is to license research output under CC licenses.

At the same time, it is important to underline that licenses allowing reuse should not only apply to contents, but to metadata too, in order to ensure their reuse for machine readability purposes.[11]

Recommendation: Encourage the ERCIM members to push their researchers for the use of Creative Commons Licences,[12] in particular of the CC-BY. If authors feel they need a more restrictive license for their work, their decision should be carefully motivated. In any case, the licenses applied should fulfill the Berlin Declaration principles.

Towards an exhaustive coverage of online publication within ERCIM members

The experience gained within various ERCIM organisations shows that even with a deposit mandate, one does not necessarily reach 100% coverage of full text deposits. There is a need to better understand the reluctances from our researchers and also provide guidelines as to the type of publication accompanying material that can be deposited.

Recommendation: Give some general guidelines as to the documents that could be deposited within publication repositories: PDF, source files, complementary material (illustrations, slides, videos).

Recommendation: Organise a simple survey towards researchers in order to better understand their reluctance regarding full-text deposits.

Repositories at the service of the publication lifecycle

Publication is the act of making public an information set. Such information could be of any kind, like texts, photos, videos, experimental data, programs in source or executable forms, etc. and could be made available on many different kind of supports like books, journals, tweets, emails, hand-written manuscripts, videotapes, blue ray discs, TV, posters, dazibao, graffitis, etc.

The purpose of a publication repository is to provide a means to store, make available and preserve forever a given category of publications, typically in our case the ones that are useful for science. A repository is open access when all its publications are. Some restrictions related for example to confidentiality (a referee report for example) or security (for instance access to documents about biological or digital viruses) may apply and should be ensured by the repository.

Digitalization has allowed for dramatic changes in repository design and use as it makes them potentially universally accessible via the Internet and digital tools.

---

[10] CC 4.0 v has been released at the end of 2013.
[11] E.g. metadata harvesting.
[12] https://creativecommons.org/choose.

Consequently, their design should allow them to be surrounded by powerful services like data mining, classification, easy versioning and analyses.

The quality of a repository is of course essential and relies on
- the quality of the data and metadata. Typical examples are:
    - the intrinsic scientific quality of the document (e.g. no plagiarism, no revisionism, etc.), which is different for the usual reviewing process of journal or conference documents;
    - the validation of the authors identification and affiliations;
- the qualities of accessibility, i.e quality of the internet access, availability and efficiency of services.

The intellectual properties questions are typically handled by adding to the document's metadata the kind of licence attached to it.

Especially when applied to scientific publications, open access repositories have an impact on the scientific approach itself and can be used not only to publish, but also to discuss or qualify the contents of open access documents. As a discussion about a document itself becomes a publication, it allows for analysing and commenting discussions in an unbounded tower of meta-discussions.

Two main well-known scientific repositories are arXiv, initiated and maintained by Cornell University and HAL, initiated by CNRS and now a joint initiative of CNRS, Inria and the University of Lyon.

Recommendation: ERCIM members should contribute to the setting up of common standards concerning publication repositories and their usage at various stages of the publication cycle.

Usage information

Publishing in the context of scholarly communication serves three main purposes: (i) legitimization, (ii) dissemination, and (iii) access, preservation, and curation (Borgman 2007).

In a world where the publishing of "the same" publication can occur in many "places", e.g. co-authors of a paper belonging to different institutions can publish for instance in their own Institutional Repositories, it is fundamental to reconcile the three purposes - initially tough in a "single place" settings - with the "many places" settings. In essence, the "many places" settings lead to a proliferation of the "instances" of the same "publication" across systems that are not necessarily conceived to know each other and to cooperate/interoperate towards a holistic and integrated scholarly communication goal. Different systems can offer diverse approaches for the three purposes.

For *legitimization*, it can be achieved by peer-review as well as by counting citations and by collecting various visibility indicators.[13] Peer-review is a well-known and consolidated process, however it is not immune from biases and new forms of peer-review are emerging, e.g. open peer-review. Citations and metrics such as usage (e.g., page views, downloads) and social metrics (e.g., Twitter, Facebook) are of limited value if not collected in a systematic and global manner.

---

[13] Cf. for instance http://www.altmetric.com.

For ***dissemination***, it is desirable that each instance of the publication is provided with links to the other instances so that every instance is actually an access point to the entire "record" of the publication that has been made available "on the web" including the files format(s). The fact that, at a certain point in time, a given instance of the "same" publication is gathering large usage statistics is a potential indicator of the "quality" of such specific instance (and the repository hosting it).

For ***access, preservation, and curation***, these facilities are usually associated with the "system" provider role. It is a duty of the "system" provider to guarantee "present" and "future" availability of the published publication.

Recommendation: Every "system" willing to expose "open access literature" should also provide for (i) the systematic collection of article-level "usage" statistics and (ii) the "open" dissemination of the collected statistics (included the metrics used and the measurement methods). Statistics should be available for each single publication. In order to favour interoperability, the exploitations of standards like SUSHI and COUNTER is encouraged.

Recommendation: Every "system" willing to expose "open access literature" should also provide its users with comprehensive information/metadata about "each" publication, including link(s) to actually access the publication "content", links to instances in other repositories, and local and global (aggregated) statistics.

The future of repositories

Beyond their current role in supporting open access to publications, we anticipate that repositories will play a growing role in the future for the management, preservation and communication of research assets at large. We can outline the main possible evolution direction as follows (Assante et al. 2015):

- covering more data types beyond publications, in particular image, audio and video assets, but also allow step by step the hosting of various kinds of data sets and software;
- allowing re-publication by departing from current fixed document formats such as PDF and having standardised XML-based representation as the default back-office representation;[14]
- enhancing services by deploying annotation functionalities or text and data mining facilities, or the management of usage indicators;
- connecting repositories to environments dedicated to community feedback or peer review (overlay journals).

In order to achieve this, we consider it necessary to go towards a **better coordination between existing platforms** reducing the current level of fragmentation that can be observed at European level. We think that ERCIM can play an essential role in facilitating concertation among their members to create a strong network of sustainable repositories in the Computer Science domain.

---

[14] See Romary and Lopez 2014 http://ercim-news.ercim.eu/en100/r-i/grobid-information-extraction-from-scientific-publications.

References

Assante, M., L. Candela, P. Castelli, P. Manghi, and P. Pagano. 2015. "Science 2.0 Repositories: Time for a Change in Scholarly Communication." *D-Lib Magazine* 21 (1/2). doi:10.1045/january2015-assante.

Bardi, P., and P. Manghi. 2014. "Enhanced Publications: Data Models and Information Systems." *LIBER Quarterly* 23 (4).

Berthaud, C., L. Capelli, J. Gustedt, C. Kirchner, K. Loiseau, A. Magron, M. Medves, A. Monteil, G. Rivérieux, and L. Romary. 2014. "EPISCIENCES - An Overlay Publication Platform." In *ELPUB2014 - International Conference on Electronic Publishing*, 78–87. IOS Press. http://www.ebooks.iospress.nl/publication/36552. 10.3233/978-1-61499-409-1-78.

Borgman, C. 2007. *Scholarship in the Digital Age.* Cambridge, Mass: The MIT Press.

Candela, L., D. Castelli, P. Manghi, and A. Tani. 2015. "Data Journals: A Survey." *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23358.

Guibault, L., and Wiebe, eds. 2013. *Safe to Be Open: Study on the Protection of Research Data and Recommendation for Access and Usage.* University of Göttingen Press.

McKiernan, G. 2000. "arXiv.org: The Los Alamos National Laboratory E-print Server." *International Journal on Grey Literature* 1 (3): 127–38.

Panitch, J.M., and S. Michalak. 2005. "The Serials Crisis. A White Paper for the UNC-Chapel Hill Scholarly Communications Convocation." http://www.unc.edu/scholcomdig/whitepapers/panitch-michalak.html.

Romary, L., and C. Armbruster. 2010. "Beyond Institutional Repositories." *International Journal of Digital Library Systems* 1 (1): 44–61. hal-00399881.