

End of Project Statements by the PEER Executive Partners

Reflections on Open Access Scenarios

International Association of Science, Technical and Medical Publishers (STM)

Göttingen State and University Library (SUB)

Max Planck Digital Library (MPDL)

Inria (Institut National de Recherche en Informatique et en Automatique)

Points of Agreement

A note from the European Science Foundation (ESF)

PEER End of Project Statement by STM

Change & Continuity in evaluating Green Open Access

a) STM and its members support sustainable open access, i.e. routes to open access that ensure the integrity and permanence of the scholarly record via a sustainable business model see www.stm-assoc.org/publishers-support-sustainable-open-access . Open access publishing via a pay-to-publish model, as offered by many STM members, is one such route.

b) Individual publishers have varying attitudes towards Green Open Access as well as a variety of policies and embargoes – but mandatory Green Open Access policies are problematic for virtually all publishers. Generally, publishers do not regard mandatory Green OA as a sustainable route because it assumes that the free availability of the peer reviewed manuscript of a paper, even after an embargo after publication, will do no harm to whatever model (if any) might be in use. It has no inherent business model to support it per se and may easily undermine any subscription-based models.

c) PEER was an effort to examine the effects of a Green OA policy at the European level and to provide some data about this type of open access that is sorely lacking. For example, embargo periods of six months continue to be adopted for mandates without any hard evidence as to whether that embargo really protects the journal from usage loss and ultimately cancellation, and how this varies by subject area and journal type. STM proposed to end fruitless hypothetical debate by seeking evidence to inform opinions and policy. This required a collaborative effort of publishers, repositories and the research community.

PEER Project achievements

a) The collaboration between publishers, research funders and the repository community has established greater trust and mutual understanding among key players, which is essential to monitoring the effects of open access policies. It rapidly became clear that involvement in the running of a complex, large-scale research project required partners to concentrate on practical collaborative issues that squeezed out politics and allowed mutual professional respect to become established

b) Because the experiment was conducted on such a large-scale (international, many leading publishers, repository community), the results (infrastructure and observatory) are not only novel (e.g. comparison of usage at publishers / repositories) but also more robust and reliable (e.g. duration / repetition of behavioural research) than any work hitherto conducted.

c) The project has delivered on all its original aims: collaborative infrastructure, tens of thousands of manuscripts, research outcomes.

d) It has demonstrated how complex it is to deliver a working Green Open Access scenario even for an experiment where all parties are supportive of the research, in terms of the entities managed, missing metadata, manipulation and transport of ill-defined file objects, and adaptation of work flows. Such voluntary collaboration should not be expected in a real world situation where additional time and cost factors might make Gold Open Access a better outcome.

The experience of building the PEER infrastructure

- a) Building an infrastructure on an international scale, even if as an experiment only, has been a cumbersome process. Publishers and repositories are all configured differently. Capturing the manuscript, transferring them to repositories, managing embargoes, ensuring the visibility of the content – all proved to be more difficult than the project partners had anticipated. It has become clear that making a Green Open Access world a reality, even if only in the limited sense of the PEER Observatory, is much more complex and involves levels of collaboration and hands on work that would be unlikely in a real world situation, and this raises questions about viability. If doing Green is cumbersome, labour-intensive and certainly not free, and if that results in a peer reviewed manuscript some months after publication, it may be better to invest in paying for the final version immediately (Gold Open Access).
- b) The experiment has been successful, but it should not be assumed that publishers would collaborate voluntarily and pro bono in future / further implementations of Green OA policy.
- c) The so-called Stage Two Version in scholarly publishing, i.e. the author's peer reviewed and final manuscript, remains ill-characterised for the authors, publishers' staff and many others. The Stage Two Version is not part of the usual workflow in scholarly publishing, and thus each file has to be vetted manually (e.g. for any comments, especially by peer reviewers, typos, corrections). Manual vetting introduces cost – making Green OA an expensive parallel world to the standard literature.

Insights from PEER Research results

- a) Usage: In a short period of exposure (just over a year) usage of the articles that are part of PEER have divided 8:1 between the official journal version and the repository one from a pre-PEER starting point of 100% of uses being with the journal. This figure may be interpreted in a number of ways, but it raises a number of questions
 - i) If the ratio of 8:1 has been reached so quickly what would the ratio be a further one or two years down the road? At what point does the shift of use to free repository versions make journal purchase uneconomic?
 - ii) Can such low levels of comparative use of repository versions justify the complications and expense of building the parallel Green open Access publishing infrastructure scholars would demand?
 - iii) Would National licensing deals (that would essentially mimic open access for the citizen) not be more cost effective for the European public?
 - iv) Would promoting and paying for Gold Open Access options yield a quicker and higher utility outcome, given that this makes the final Version of Record article available immediately upon publication?

Moreover, it must be observed that by its nature Green OA lives off subscription-based publishing (by title, in bundles) and the parallel availability of Green OA versions principally inhibits the development of other access models by

- i) Usage-based pricing
- ii) Pay-per-download

as well as potentially undermining the subscription model itself with destructive effects to the journal.

- b) Behaviour: Overall, the results show that scholars overwhelmingly prefer a gradual evolution of the publishing system, with an emphasis on trusted journals and the version of record. Unsurprisingly, researchers primarily do not associate open access with self-archiving and consider repositories as complementary only. For scholarly work, access to the version of record is essential. In detail, the research indicates that
 - i) Self-archiving is not prevalent among authors, nor considered a duty; of c.11,000 authors contacted by multiple invitations to deposit from the publisher only 170 actually did so;
 - ii) Readers have concerns about the authority of article content and the extent to which it can be cited when the version they have accessed is not the published final version.

- c) Economics: The study on cost structure of publishers and repositories is based not on models and estimates but on data gathered, much of it confidential business data. The case studies show that cost structures are disparate. In particular the study notes that
 - i) The costs for managing just the peer review are not negligible (USD 250 per submission) and that no economies of scale apply;
 - ii) The cost of publishing and platform maintenance varies, but it is also real, and a necessary investment to ensure visibility and access.

The study indicates that ongoing investment in digital platforms is required – for a number of reasons, e.g. manage content growth, added information services, archiving, technological innovation. The study notes that repositories often are characterized by a high level of sunk costs, at a time when public resources are expected to become scarcer with austerity cuts. Publishers are much more likely to make the investments necessary to drive forward scholarly communication.

Outlook: What future(s) exist for Green OA?

Taking all the evidence of the PEER research and the experience of building the Observatory together, rather questions the wisdom of investing huge amounts of time and public money in the pursuit of a Green Open Access parallel universe. There are important roles for digital repositories to play in the preservation of the digital outputs of scholarship and the vexed research data issues. At the level of practicality, user acceptance and utility of outcome, the Green approach seems much less attractive than a Gold pay-to publish one: Gold creates no risks for the journal model, it allows scholars and the general public to have the final versions they really want upon publication, it involves no additional infrastructure developments (such as PEER had to create with its Depot) or assuming (unlikely) publisher and repository collaboration, and would free up money and staff at the repositories to concentrate on data archiving that is at present extremely poorly dealt with.

PEER End of Project Statement by UGOE

Change & Continuity in evaluating Green Open Access

Since 2005 the University of Göttingen calls their scientists upon to publish Open Access. This is a strong recommendation instead of a mandate. In 2007 the University's Foundation Council underlined the need to further implement Open Access, which was echoed 2009 by the Deans of the Faculties of Medicine, Natural and Life Sciences, Mathematics and Informatics.

The Göttingen State and University Library (SUB) provides both green and gold OA support structures for authors of the University. Infrastructures like GoeScholar, the institutional repository for peer-reviewed publications, or Göttingen University Press for Open Access book publishing is backed up with several information or library services. Starting in 2004 the Library has negotiated an increasing number of institutional agreements with OA publishers, adding up to 8 contracts in 2012 (with further plans for the future). In addition, the Library has a special (pilot) agreement with Springer SBM since 2007. Most of the agreements allow a deposit of stage-3 publisher's versions of articles and metadata into GoeScholar. Gold and green OA are converging at Göttingen University, as those OA publications have a major share in the GoeScholar repository.

The University of Göttingen considers Open Access to be an essential publishing strategy for publication and research data. The repository today is not only considered to provide open access to publications but to become a permanent node as content provider in a digital, globalising knowledge infrastructure network (cf. OpenAIRE).

PEER Project achievements

Major achievements of the PEER project include:

- PEER is a pioneering, to date unique project in terms of the involvement of controversial stakeholders.
- The project has implemented in vivo a large-scale deposition process of articles from publishers into repositories.
- Publishers provided feeds for 241 heterogeneous journals and over 200 control journals
- A technical infrastructure has been set up and practise *Guidelines for publishers and repository managers on deposit, assisted deposit and self-archiving* <http://www.peerproject.eu/reports/> have been released which can be used to implement in production mode the article transfer from publishers to repositories beyond the project lifetime.
- More than 10,000 research articles can be offered to the scientific community and the public by partner repositories beyond the project's lifetime.
- Implementation of the SWORD protocol to allow application-level deposit of material into repositories.
- The development of the PEER Depot as a central intermediate clearing house, receiving content from publishers and distributing it to repositories. The PEER Depot developed among others the following tools:
 - Special technical adaptations tailored to the needs of various stakeholders: Extraction of additional metadata from PDFs; additional filtering of social sciences content for ingest in a participating subject repository
 - An embargo management facility at PEER Depot level to manage the different embargo periods assigned to each journal
 - A unique exchange format of metadata (publishers / repositories) by means of a TEI customisation plus the mapping of different metadata schemas (e.g. NLM and proprietary schemas)
- PEER is the only Green OA project accompanied by systematic collection of data on behaviour, usage & economics
 - Behaviour: Unique collaboration with publishers and authors to reach out to authors and users

- Usage: First large-scale and comparative collection of article level usage statistics
- Economics: First detailed empirical study of cost drivers of publishers and repositories

The experience of building the PEER infrastructure

- The construction of the PEER Depot as central intermediate deposit and distribution instance has been instrumental to facilitate the collaboration of publishers and repositories at an early stage. Its set-up is one of the major achievements of the project. The PEER Depot is a facility that handles heterogeneous data formats, detects duplicates and manages embargo periods. This central instance where content is received, held under embargo in a dark archive, matched with corresponding metadata and distributed to repositories participating in the project holds potential for applications beyond and outside the project. In the future further investigation is necessary to find out, whether the PEER infrastructure architecture can be complemented by direct content transfer from publishers to institutional and disciplinary repositories.

– One major advantage for repositories related to article integration is the implementation of the SWORD protocol as deposit mechanism. The SWORD interface, once implemented, has applications beyond the project and may be adopted by repositories to accept material directly from various sources, e.g. publishers, in the future. Also, it allows for interoperability between different repositories.

– The implementation of the PEER Guidelines, created along the lines of the DRIVER Guidelines (today OpenAIRE), resulted in a standardised transfer, dissemination and acceptance process which is also applicable beyond and outside the project.

– Some challenges WP2/3 were facing:

- Number of articles authored by EU-located researchers was lower than expected.
- Defective stage-2 articles received from publishers. As articles are not checked for contents at PEER depot level, some defective articles, holding for example confidential reviewer comments, were distributed to repositories which resulted in considerable time and effort for detection and removal.
- Number of repositories willing to join the Repository Task Force was lower than expected. The project encountered political as well as sociological and technical challenges for repositories when invited to join the Repository Task Force.

– Some challenges repositories were facing:

- Individual adaptation to PEER workflow needed for each repository: implementation of SWORD protocol, TEI mapping, log file transfer to usage research team etc.
- Additional adaptations to PEER depot workflow were necessary during the first months of article transfer and also current monitoring for requested subsequent adaptations.

– Some delays within the project were due to technological changes. These challenges could not have been foreseen at earlier stages of the project and are mainly due to its complexity.

Insights from PEER Research results

In general the findings of the Behavioural Research Team outlined in two reports, were not unexpected. Based on these studies several common-knowledge assumptions such as a low compliance rate for authors to self-archive can now be considered as confirmed and validated by empirical data.

The results of the Usage Research also nail down common assumptions. We can assume that the PEER experiment has not impacted publishers negatively and that repository use is possibly complementary use. What is more, Usage Research has shown that there is enhanced traffic at publishers' websites through green OA versions being available at repositories.

Insights from the Economic Research were less strong as expected. This can possibly be traced back to the fact that analysing a field as complex and unexplored as systematic green Open Access would have required to allocate more project time and resources to produce robust findings. Future research would need to highlight the true costs associated with archiving stage-2 articles under different business models and use cases.

Outlook: What future(s) exist for Green Open Access?

In terms of Green OA the PEER behavioural research confirms that copyright issues and publishers' self-archiving policies are perceived as a major barrier to self-archiving. Generally, authors do not know whether they have permission to upload a copy of their article into a repository or website, nor do they know which version (pre-print, author's final manuscript or publisher's version) they are entitled to deposit. We consider these aspects the driving force for author's remarkably low compliance rate in self-archiving their articles. Publisher-mediated deposit or publishers informing authors about their deposition rights and which version they may deposit in which open access repository, will be of considerable help in overcoming authors' concerns.

Mandatory author deposits might also encourage authors to self-deposit, since in recent years e.g. universities as employers of academic staff and funding agencies in most disciplines have developed open access policies or at least position statements on open access. Furthermore, where funding agencies stipulate mandatory deposits they typically also provide financial support for author-pays mechanisms.

It is obvious that appropriate action from publishers can significantly stimulate Green OA, i.e. populate repositories. In the PEER experiment, publishers' deposits provided the vast majority of articles for participating repositories compared to authors' self-deposit which only contributed a negligible number of articles.

The PEER experience shows that it is crucial to convey positive messages and incentives for authors to self-archive their manuscripts: *What is in it for them if they self-deposit?* In addition, it is necessary to find creative strategies by using new technologies: *What tools can be used to facilitate the process?* In short, self-deposit is perhaps a good starting point for green OA scenarios, but a combination of strategies seems advisable, in particular to obtain the version of record, reliable metadata and a regular flow of articles.

In practise, gold and green OA are increasingly converging, as gold OA publications provide the easiest and most comfortable way for authors and repository managers to fill repositories with content. Different from the initial set-up of repositories, their mission has been significantly expanded from a local open access portal and publication bibliography to an essential building block of the upcoming e-Infrastructure. Repositories today are considered to be nodes in the future, global knowledge infrastructure network, where thousands of publication, data, learning material and other content repositories provide a knowledge base which can be accessed and (re)used in seamless ways by researchers and the interested public. Networks of repositories such as COAR are facilitating this process.

PEER End of Project Statement by MPDL

Change & Continuity in evaluating Green Open Access

The Max Planck Society has no formal mandate enacted but strongly recommends following the principles of Open Access. In addition to publishing or supporting several Open Access Journals (Living Reviews Journal Family; eLife), the Max Planck Society maintains a central institutional repository, which is responsible for capturing the Society's research output and acts as a key component in the emerging eResearch infrastructures as well as an Open Access instrument. At the same time, the Max Planck Society is also very active in Gold Open Access, maintaining a substantial and growing number of publication cost agreements with major Open Access publishers. With international partners in the research communities and in the publishing industry, the Max Planck Society is preparing infrastructures that would allow for a sustainable transformation of existing journals from their current subscription basis to a publication cost model. The Max Planck Society strongly believes in a publication cost future for scholarly communication but also adheres to the important role of repositories as an institutional node in the distributed and shared eResearch infrastructures. In that respect, PEER has helped to gain experience and further understanding of workflows and challenges but has not much altered our fundamental points of view.

The PEER Project achievements

The PEER project can demonstrate a number of major achievements, among them:

- Involvement of non-Open Access (OA) publishers in the provision of research articles (241 journals and 200 control journals)
- Well defined depositing workflows and processes for publishers, authors, PEER Depot and participating Repositories (see *Final report on the provision of usage data and manuscript deposit procedures for publishers and repository managers* and *Guidelines for publishers and repository managers on deposit, assisted deposit and self-archiving*, available at <http://www.peerproject.eu/reports/>)
- Developed accompanying technical infrastructure to support the deposits of research articles from publishers or participating authors and their redistribution to multiple OA Repositories, comprising of:
 - Central PEER Depot which holds descriptive and administrative article metadata
 - Tooling to extract metadata from article PDFs and their transformation into a unified TEI-based metadata exchange format
 - Tools to support the embargo management according to different embargo periods for different journals
 - Tools to report on deposits by publishers, authors, embargo periods and PEER Depot statistics
 - PEER Help Desk - supporting deposits of manuscripts by authors and providing help-desk features for interested parties
 - Long-term preservation of all articles at the KoninklijkeBibliotheek Netherlands

- More than 18,000 research articles can be offered to scholars and general public by each participating repository
- Studies on behaviour, usage and economics through publisher and repository participation, supported through mechanisms such as publisher communication to authors, repository surveys and usage logs from repositories

The experience of building the PEER infrastructure

Several activities had to precede the actual implementation of the PEER infrastructure: an analysis and investigation of possible workflows and mechanisms, which can be adopted by both publishers and repositories. The PEER Depot is a central "dark" archive, which holds article deposits from publishers until their embargo period expires. After this period, articles are deposited to the participating repositories. Despite the complexity of interactions, two simple mechanisms were selected to deposit articles to repositories: SWORD enabled server interfaces and FTPS.

Besides valuable domain and technical experiences of partners involved in the project, several factors largely contributed to the success of the whole process, thus resulting in substantial amount of new content in repositories.

Primarily, definition and adoption of single metadata format (along the DRIVER Guidelines) from the very start of the project had been done at a very early stage of the project. TEI formatted metadata and the quality of the automated extraction from the publisher PDFs enabled for early stabilization of the transformation procedures that had to be implemented on repository sides.

The internal tooling implemented on the side of the PEER Depot allowed for detection of duplicates, metadata matching, management of embargo periods and selection of relevant articles already on the PEER Depot side, thus the articles deposited to participating repositories were validated (on metadata level) for eligibility before their depositing to the repositories.

The adoption of the SWORD as depositing (push) protocol between the PEER Depot and the repositories was preferred over OAI-PMH (pull) protocol. Although this required certain efforts by repositories, clearly defined interfaces and enabled client libraries in several programming languages enabled relatively smooth adoption and integration in the repositories. As the protocol enabled repository responses about the success/failure of each particular deposit to be sent back to the PEER Depot, reaction time to eventual depositing failures was minimized. This protocol was additionally used to distribute articles deposited manually by authors via PEER Help Desk, after metadata were matched in the PEER Depot and embargo period expired.

Delays existed in the delivery of functioning PEER infrastructure due to technical interventions on PEER Depot side or during adoption by participating repositories. However, from MPDL aspect these were not seen as critical.

Challenges emerging during infrastructure development, with significance beyond the PEER project:

- How to enable content (i.e. PDF) validation - sometimes articles containing only front page, confidential reviewer comments were distributed to the repositories. Detection was not automated, therefore it caused extra efforts to remove them from repositories and eventually redistribute them with valid content

- How to integrate the PEER content with the regular repository content, considering the quality standards of the repositories – several approaches were taken in this respect: some repositories set-up separate PEER repository, while others integrated the PEER content directly in the repository. In case of SSOAR, for example, each PEER article underwent person-driven quality assurance process. This diversity however affected the results in the analysis of the usage data provided by each repository.
- A question to be answered is what happens to the content beyond the PEER Project duration. Different repositories may have different strategies; however this decision may be influenced by the level of integration of the PEER content with the regular institutional content.
- What happens to the PEER Depot infrastructure and could it be useful for feeding content to repositories in future?

Insights from the research

The three research studies were confronted with extremely high expectations and with a very complex project setting that had to overcome many fundamental technical and organizational challenges, thus leading to delays and unanticipated challenges for the research groups as well. As often when expectations are inflated, the moment of delivery brings a certain soberness that may come as a surprise at first but then, after a moment of reflection, can be seen as the reality that could have been expected. The PEER research results did neither change the world nor necessarily the outlook to the world; but they brought substantiation and analysis to areas which had only limited evidence so far: Behavioural research gathered empirically data from surveys running at all participating repositories and by direct work with smaller set of researchers via focus groups and participatory workshop. The results can be characterized as additional empirical confirmation of more or less expected facts. The economics report listed and discussed various cost structures with many interesting aspects, details and analyses; but no new picture really emerged.

Outlook: What future(s) for Green OA?

Apart from confirmed copyright issues and self-archiving policies of publishers, no clear understanding emerges what authors gain if they self-archive. Exceptions are certainly some communities e.g. arXiv, where researchers mostly discuss over pre-prints, therefore this actually is natively integrated within their scientific workflows.

General experiences from PEER project show that author deposit (even though encouraged and mediated by publishers) was minor (almost negligible) with respect to the overall number of articles. Most of the content resulted by the publisher deposits. According to the PEER behavioural analysis, many of the answers about awareness of OA repository or archiving of content within the institutional repository were practically answered as "someone did it for me".

Another aspect to consider is whether/how useful would be the Green OA approach for researchers in an atmosphere where no particular issues arise around the visibility of the research, or the access to journal articles, unless mandated from the funding institution. In this case, Gold OA publications would generally be more attractive, especially as they do count in scientific research impact metrics.

Answering the question of "why to self-deposit and how it helps me" is crucial for further understanding the future potential of Green OA. Approaches vary from mandatory deposits

through support by local libraries and publisher mediated deposit. They need to be carefully investigated and supported by appropriate technological platforms. Such Green OA enabling platforms must strongly contribute in bridging sensitive issues such as metrics and research impact. Consequently, their offer should be expanded with attractive functionality such as semantic searches, possibility to download publication data sets, smooth integration with research data, rich publication and presentation experience for the end user, collaborative features etc., and with a variety of additional services that span the borders of a single Green OA enabling platform and pure data dissemination.

PEER End of Project Statement by Inria

A technical infrastructure beyond ‘Green’ and ‘Gold’

PEER – a successful project for Inria

Ever since the PEER Project began to take shape in May 2007, Inria has been interested in understanding the technical implications of a large-scale deposit process. Once the PEER Project launched in the autumn 2008, the PEER Depot was defined and implemented. This powerful platform can ingest metadata and documents from a variety of publishers’ proprietary formats and distribute it further to a whole network of publication repositories. This allows us to think beyond the quite dated categories of ‘green’ and ‘gold’ open access.

Indeed, through our interaction with publishers and publication repositories, and the creation of an interoperability platform between them, we have identified that there is a whole range of possible scientific information scenarios where there could be a need to synchronise the information held by public and private actors in the scholarly publication workflow. Such scenarios may result from a funder’s deposit mandate, a specific agreement between a research organisation and a publisher, but also more globally on a joint understanding that whatever the assumption about open access, there is some information that for reason of public interest should freely circulate among scientific information actors.

In this respect, the PEER infrastructure (Depot) opens up the possibility of creating a European (or even global) clearinghouse for the exchange of publication data (metadata as well as full-text) that can be further tuned by the corresponding actors. For Inria, whose strategy in the domain of scientific information is to command its scientific information patrimony and achieve maximal impact for its research results, the PEER technical platform has become an essential element in the design and further orientation of its scientific information policy.

General background on the scientific information policy of Inria

Inria has been involved in the open access movement for many years. It was an early signatory of the Berlin declaration (2003), organized the 7th Berlin conference in Paris in 2009, and participated in various working groups on open access and scientific information.

At a national level, Inria has always worked closely with the CNRS on the evolution of the national publication archive infrastructure HAL, which is, since 2005, the unique deposit platform recommended by Inria. Furthermore, Inria has been a leader in the development of a national agreement between universities and research institution on the systematic deposit of publications on HAL. Inria has also joined the national endeavour to coordinate scientific information policies and infrastructures in France, BSN (Bibliothèque Scientifique Numérique), particularly in its Open Access and Scientific Publication sub-committees.

At a European level, through a membership to Couperin (consortium of research and high education libraries) Inria is participating in the activities of the OpenAire* European project(s).

Beyond the (soft) deposit mandate implemented within Inria, which has led to a very high coverage, Inria has experimented with supporting native (or Gold) open access by both

allocating a dedicated budget to researchers and entering into global negotiations with native open access publishers.

All in all, continuity prevails in Inria's vision and behaviour regarding Open Access, but changes occurred in the implementation of green and gold OA from an editorial and technical point of view.

Challenges and achievements in developing the PEER Depot

The main task of Inria in the PEER project was the set-up and exploitation of the PEER Depot, the technical platform that gathers, filters and disseminates metadata and manuscripts from publishers to publication repositories.

The main challenge while building the PEER Depot was to create the necessary infrastructural processes and protocols for depositing and distributing articles to participating repositories (HAL being one of them). The variety of partners involved in this task made this task a particularly complex one: 6 participating repositories, 12 publishers involved, 241 journals providing articles. Another technical challenge was the ability to filter EU content, embargo period, subject field and article type.

A core activity related to the implementation of the PEER depot has been to work on a unified data structure that would normalize, and at times complement the metadata provided by the publishers. This metadata curation environment (quality control, embargo management), based in particular on the TEI¹ guidelines is an essential component which will be reused internally, for instance in the context of the evolution of the HAL platform.

Experience gained through participation in the PEER project

The quality of exchanges among partners has been high for the entire project and it led to very satisfying common work with publishers: communication channels for editorial (e.g. content selection) and technical (e.g. metadata formats) issues were open and efficient.

In terms of results, this good common work led to the implementation of a technical platform covering the publishers main formats as well as the elaboration of a common (and elegant) representation which integrates metadata but also plain text. The elaboration of a uniform metadata format is one of the main achievements in PEER, especially considering the heterogeneous quality of publishers' data (e.g. levels of granularity in affiliation information).

The implementation of the PEER Guidelines made possible a satisfying workflow of data in and out of partner repositories.

The PEER project also offered the opportunity to test an emergency workflow enabling the automatic extraction of metadata from PDFs. The capacity to automatically extract metadata from the manuscript was a groundbreaking technical development.

The prospect of the PEER observatory

Inria scientific and technical information staff, which contributed to the research carried out within the PEER project, was especially involved in the economics research. It gave feedback about investments in online repositories and the associated costs. The finding of the Economics Research did not significantly diverge from the point-of-view of Inria. The main difference results from the centralised nature of HAL as a national platform. As it is shared

¹ Text Encoding Initiative

with other French research institutions, its usage costs are very low compared to the number of papers or researchers.

Even if the involvement of Inria on research aspects was not very high, we would support to make the experimentation platform more sustainable in the form of we have thought of, back in 2007, as the PEER observatory.

Further scenarios

The technical experience raked in during the setup of the PEER Depot makes us aware of the possibility of using such an infrastructure as a data integration platform adapted to various OA scenarios. In case of an agreement with publishers, it would be possible to deposit stage-3-articles and use the PEER Depot as a clearinghouse for the distribution of OA content towards repositories.

INRIA is currently defining its strategic plan for the scientific and technical information. All PEER results and research documents will be spread within the scientific and technical information network so that it contributes to the elaboration of this future strategy.

Do we need more colours than Gold and Green?

The PEER experience suggests that more ambitious scenarios should be pursued, such as the setup of virtual research environments, where scientific information flows seamlessly between researchers and certification entities: “In this research workspace, the scholar manages all the stages of his scientific information activity, gathering initial evidence by importing data from external repositories, on the form of “publications” or as extractions of existing observations or documents provided by other scholars. Building up from this evidence, he compiles and organises his thoughts in the form of drafts or annotations directly linking to the other documents in his workspace, applying specific software to compute new features from his observations, and organise his thoughts by grouping meaningful sets of information in dedicated research folders.” (<http://hal.inria.fr/hal-00659856>)

This environment should be capable of integrating information sources, multiple publications, research data, etc. It would also offer scholars the possibility to select the diffusion mode most appropriate to their productions (entirely private, accessible to a defined list of persons, totally OA). That is the direction INRIA would like to go.

Related publications:

<http://hal.inria.fr/hal-00659856>

<http://hal.inria.fr/inria-00537302>

<http://hal.inria.fr/inria-00593677>

Points of Agreement

The PEER Project Partners would like to indicate that they agree on the following points:

1. Building a large-scale infrastructure is organizationally and technically challenging

When the PEER Project started, there was no European infrastructure available that was robust or scalable or efficient. Hence, with considerable effort this infrastructure was built, linking publishers and repositories to the PEER Depot as central clearing house.

2. Building a clearing-house with automated workflows is helpful

What made the PEER infrastructure a success is the ability to construct a largely automated workflow for the ingestion and distribution of articles.

3. Author self-archiving is unlikely to generate a critical mass of Green OA content.

The author deposit rate in the PEER Project was exceptionally low. This unwillingness to deposit, even when the author explicitly is invited by the publisher, suggests that author self-archiving will not generate a critical mass of Green OA content.

4. Stage II archiving requires manual oversight and intervention

The author's final peer reviewed manuscript (the so-called Stage II manuscript) remains difficult to handle for publishers, repositories, authors and readers, requiring manual oversight and intervention.

5. Scholars prefer the Version of Record

The behavioural research as well as usage log analysis indicates that scholars prefer accessing the version of record.

6. Usage scenarios for Green Open Access are more complex than generally acknowledged

While usage at repositories may be described as a percentage of usage at publishers' platforms, and, conversely, repositories have a function for users in developing countries, usage patterns on the Internet are more complex, with the PEER repositories driving usage to publisher platforms.

7. The acceptance and utility of open access publishing has increased rapidly

Open access publishing is increasingly important for publishers, repositories and the research community. Any discussion of future Green OA scenarios must take account of this development.

8. A successful collaboration for experimental results

In the Green OA debate, the PEER Project partners started from conflicting positions, and were dependent on the support of publishers and repositories, but were nevertheless able to deliver the experimental infrastructure and observatory research to a mutually satisfying conclusion.

9. Mutual understanding and trust

Working together to deliver the project - Building the infrastructure together, getting the deposit process to work and commissioning the research encouraged - particularly also in challenging or difficult moments, engendered professional respect on all sides.

A note from the European Science Foundation

The European Science Foundation (ESF) has supported the PEER Project and is satisfied with the results presented. Due to considerable changes within the ESF, we are unfortunately not able to provide an end of project statement. We are in line with the common areas of agreement that the partners have established in terms of the results of the PEER project. ESF is pleased that the project has made a valuable, evidence-based contribution to the Open Access debate, as this is an issue that is of great concern to our 72 members from 30 countries, which are made up of research funding bodies and councils.