



**HAL**  
open science

## PEER D7.1b Additional Outcomes

Chris Armbruster, Heidemarie Hanekop, Julia Wallace

► **To cite this version:**

Chris Armbruster, Heidemarie Hanekop, Julia Wallace. PEER D7.1b Additional Outcomes. [Technical Report] 2012, pp.56. hal-00737102

**HAL Id: hal-00737102**

**<https://inria.hal.science/hal-00737102>**

Submitted on 1 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## ECP-2007-DILI-537003

### PEER

### PEER: Additional Outcomes

<b>Deliverable number/name</b>	<i>D-7.1b</i>
<b>Dissemination level</b>	<i>Public</i>
<b>Delivery date</b>	<i>6 June 2012</i>
<b>Status</b>	<i>Final</i>
<b>Author(s)</b>	<i>Chris Armbruster, Heidemarie Hanekop and Julia Wallace</i>



**eContentplus**

This project is funded under the *eContentplus* programme<sup>1</sup>,  
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable

---

<sup>1</sup> OJ L 79, 24.3.2005, p. 1.

## Contents

<b>1. INTRODUCTION: ADDITIONAL OUTCOMES .....</b>	<b>3</b>
<b>2. THE PEER INFRASTRUCTURE: CHALLENGES, ACHIEVEMENTS AND REFLECTIONS.....</b>	<b>4</b>
<b>3. POINTS OF VIEW: PUBLISHERS.....</b>	<b>10</b>
BMJ JOURNALS.....	10
EDP SCIENCES.....	12
ELSEVIER .....	14
IOP PUBLISHING .....	16
NATURE PUBLISHING GROUP .....	18
OXFORD UNIVERSITY PRESS .....	20
PORTLAND PRESS.....	22
SPRINGER.....	24
TAYLOR & FRANCIS.....	26
WILEY.....	28
<b>4. PEER FROM THE PERSPECTIVE OF THE REPOSITORIES: EXPERIENCE, OUTCOME AND CONCLUSIONS FOR OPEN ACCESS POLICY .....</b>	<b>30</b>
4.1 EVALUATION OF THE PEER DEPOSIT MODEL.....	31
4.2 EVALUATION OF THE TRANSFER PROCESS AND THE TECHNICAL SOLUTIONS IMPLEMENTED BY THE REPOSITORIES.....	33
4.3 OUTCOME FOR THE PARTICIPATING REPOSITORIES AND CONCLUSIONS FOR A FUTURE OA STRATEGY .....	35
<b>5. AUTHOR DEPOSIT: NUMBERS AND REFLECTIONS .....</b>	<b>37</b>
<b>6. USER EXPERIENCE: REPOSITORY EXIT SURVEY .....</b>	<b>41</b>
<b>APPENDIX.....</b>	<b>50</b>

## 1. Introduction: Additional outcomes

The overriding aim of the PEER Project has been to investigate the effects of large-scale green open access. This has entailed building an infrastructure by involving publishers and repositories, designing an author deposit interface and enhancing the user experience.

The means whereby the PEER Project has achieved its aims have engendered further results and reflections, thus adding value to the PEER Project.

The following is a collection of these additional outcomes. Each outcome is independent. Hence, each section is best read individually. The review proceeds as follows:

- Infrastructure: A systematic description of how the PEER infrastructure was built, how the PEER Depot works, and what the challenges and achievements were;
- Publishers: An end of project statement from the participating publishers, reflecting on the project and on change and continuity in green open access scenarios;
- Repositories: An analysis of the experience of the repositories, distinguishing between the deposit model and the transfer process;
- Author deposit: A review of the deposit process and some reflections on the very low number of authors self-archiving;
- User experience: A pilot survey seeking to understand users' experience with repositories.

## 2. The PEER Infrastructure: challenges, achievements and reflections

The PEER Project was organisationally and politically complex in nature and benefitted from the sustained collaboration of all partners and participants despite any underlying differences in policy they may have regarding green open access. This approach has also fostered respect and trust between participants who represent the wider stakeholder communities involved in scholarly communication.

The pro-active collaboration of PEER participants enabled the project to create a robust and scalable infrastructure in support of a unified and standardised ingestion and distribution service. Although some manual intervention was still required in instances where content was incomplete or delivered with errors, the vast majority of processing was automated and made use of the impressive array of technical solutions implemented in support of the project infrastructure. The PEER Depot was able to process thousands of manuscripts in a day.

### *Providing content*

To provide content, participating publishers offered articles from 241 journals within four broad subject areas: life sciences, medicine, physical sciences and social sciences & humanities. Also, these journals were selected with a range of two-year Impact Factors as reported by Thomson Reuters Science (formerly ISI). A significant portion of journals with high Impact Factors were included as well as average and low Impact Factor journals - plus some journals, which did not have an Impact Factor at the time of selection. The selection process for the participating journals is outlined within the PEER website [1].

The participating journals provided EU authored manuscripts for the project either by publishers directly submitting accepted manuscripts (& metadata), or by publishers inviting authors to self-deposit their accepted manuscripts, with the publishers providing matching metadata for validation and identification purposes. From late 2009 until the end of 2010, within each of the four broad subject areas covered by the journals, 50% of eligible articles were assigned to the publisher deposit route and 50% to the author deposit route.

### *Publisher deposits*

Feeds to the PEER Depot from publishers started at the end of the first year of the project in autumn 2009. To help build a critical mass of embargo expired content as early as possible, back-content (from the selected journals) was also provided by a number of publishers in addition to the live content feeds. In many cases, the back-content files had to undergo additional processing in order to meet the file format and metadata requirements of PEER.

### *Author deposits*

The project had initially expected an author response rate of around 10-15%. However, throughout the project, PEER observed an author deposit rate under 2%. Therefore, at the end of 2010 it was decided to transfer 48 of the journals in the author deposit route over to the publisher deposit route in support of ensuring a critical mass of content was available in the repositories for the purposes of the usage research.

**PEER content in repositories**

With the exception of SSOAR, a social sciences ‘subject repository’, each of the participating PEER repositories hosted all valid PEER content. The availability of PEER content in mirror sites enhances the discoverability by search engines, while the availability of a subset of content in a subject based repository allows comparisons with usage at institutional and national repositories.

**PEER Depot**

A central facility, the PEER Depot, was created at Inria, which acted as a clearing house, processing unit and dark archive for all submitted content. Publishers provided accepted manuscripts as PDFs. Metadata was provided by publishers either in one step (on publication) or two passes (on acceptance and on publication), following the project guidelines [2].

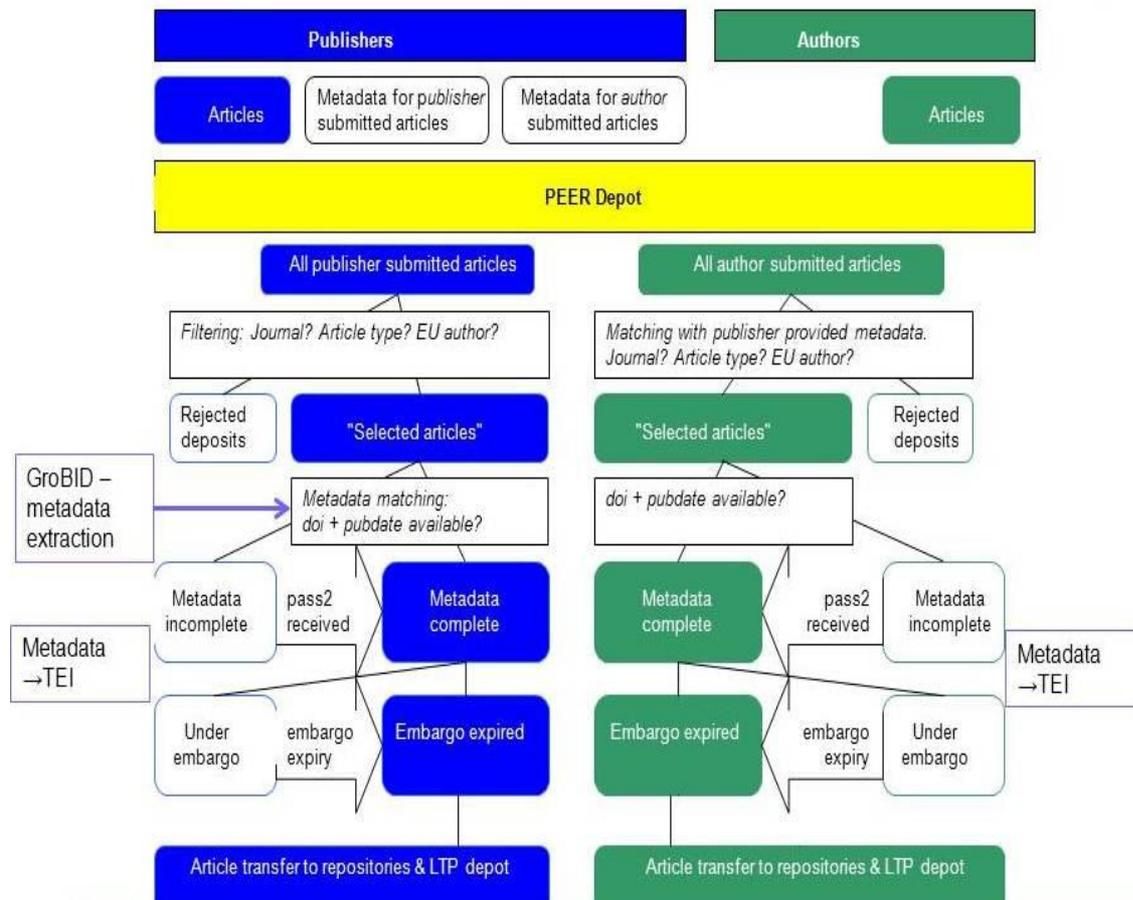


Figure 1: PEER Depot workflow

Following an invitation from the publisher, authors could submit their manuscripts as PDF files via a specially created centralised author submission interface within the PEER Helpdesk, hosted by the Max Planck Digital Library. Checks were made at this initial stage to

ensure the author had selected a valid PEER journal. The submitted article was sent directly to the PEER Depot for further processing and matching with publisher metadata.

Once content was received by the PEER Depot, publisher metadata, which was provided in a variety of schemas including NLM 2.x, NLM 3.0, ScholarOne and publisher proprietary schemas was mapped onto a single and well constrained TEI structure [3].

The content also underwent various checking procedures to ensure:

- journal validity (is it a participating PEER journal?)
- EU author
- article type (only research related content is participating)
- matching metadata is provided for each article

For the manuscripts which passed this filtering stage and were matched with metadata, further checks were undertaken to ensure all mandatory metadata elements including DOI and publication dates were present. If they were not yet available, the manuscript was held by the system awaiting the completion of metadata. If all metadata was provided, the manuscript processing was completed and the manuscript then held by the PEER Depot for the embargo period specified for that particular journal. Embargo periods for journals participating in PEER varied from 0 to 36 months. On expiry of the embargo period, manuscripts and TEI metadata were distributed to participating repositories via the SWORD protocol. The PEER Depot was identified as an authenticated source of content for each of the participating repositories. Figure 1 shows the main steps of the PEER Depot workflow [4].

In addition to the above processes, the PEER Depot developed the capability to extract metadata from PDFs using the GROBID (GeneRation Of Bibliographic Data) environment [5,6], which was trained to match various title page styles in scholarly papers. This system was used to acquire additional metadata elements, e.g. author affiliation, for over 1500 manuscripts from one of the participating PEER publishers. In principle, this process could be used to enhance the metadata record for any PDF collection of articles, which follow the style of a scholarly paper.

Figure 2 shows the content status at the end of PEER mapped onto the PEER content flow diagram, while Tables 1 and 2 show content levels at crucial intervals of the PEER Project, spanning the key period of time covered by the usage research. The tables also show the content levels at the various processing points of the PEER Depot workflow as described in Figure 1, from the initial manuscript deposit onwards, including the filtering by the PEER Depot for EU corresponding authors. Also noted is how many manuscripts had passed their allocated embargo expiry date and had been sent to the participating repositories. The non-EU, non-research manuscripts, which were filtered out were held in the dark archive of the PEER Depot and did not actively participate further in PEER, which focused on EU authored research content.

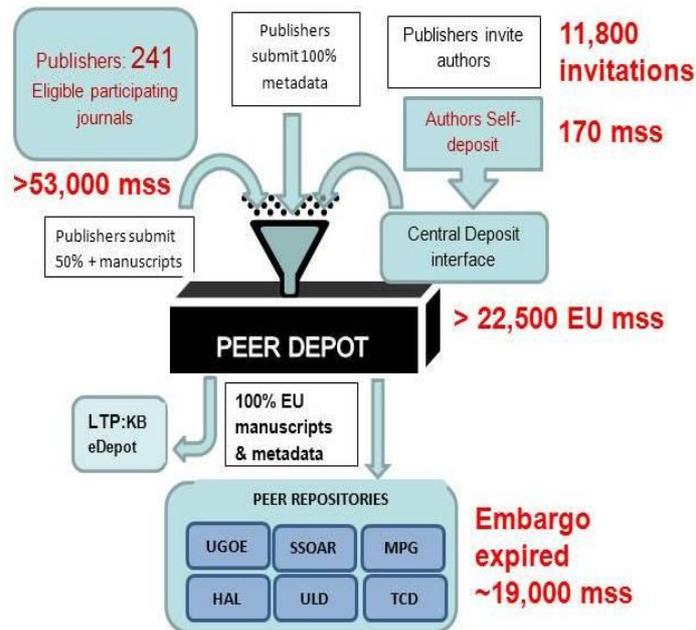


Figure 2: Content levels

Table 1: PEER content status at 01 March 2011

Type of Deposit	Total number	EU deposits	EU deposits – incomplete (1,2)	EU deposits – processed, under embargo (3)	EU deposits – embargo expired (3)	Rejected deposits (4)
Publisher deposit	39251	17116	1763	4194	11159	22135
Author manuscript deposit	125	N/A	6	27	12	0

Table 2: PEER content status at 28 March 2012

Type of Deposit	Total number	EU deposits	EU deposits – incomplete (1,2)	EU deposits – processed, under embargo (3)	EU deposits – embargo expired (3)	Rejected deposits (4)
Publisher deposit	53353	22498	1838	2365	18295	30855
Author manuscript deposit	170	N/A	19	28	52	24

Legend:

1- 'Incomplete' deposits are awaiting metadata elements e.g. DOI, publication date 2- For Author 'Manuscript Deposits', 'incomplete' means articles are matched and are awaiting additional publisher provided metadata 3- For Author 'Manuscript Deposits' only manuscripts matched with publisher provided metadata can be processed 4- Rejections are due to: i) unknown journal title, ii) non-EU authors or iii) is not valid article type

Additional explanation:

a) Total number of deposits = EU deposits + Rejected deposits b) EU Deposits = EU Deposits-incomplete + EU Deposits-processed + EU Deposits - embargo expired

### ***Technical infrastructure challenges & sustainability***

In the early stages of PEER, project partners realised that there were many challenges involved with the transferring of content from publishers or authors directly to repositories in a systematic and scalable way, including:

- Non uniformity of publisher outputs at acceptance stage (file formats / metadata schemas/ metadata elements)
- Varying requirements by repositories (file formats / metadata schemas / metadata elements)
- EU & article type filtering of content
- Embargo management at repositories
- Author authentication for deposit (ensuring authors from outside an institution could deposit)
- Non uniformity of log files
- Format problems with back-content files

Following extensive consultation with PEER publishers and repositories, solutions were found for all of the above challenges. PEER has successfully developed a robust project infrastructure, which has processed over 53,000 manuscripts and has successfully linked feeds from 12 heterogeneous publishers, and 170 authors (but in principle an infinite number of authors) to 6 heterogeneous repositories.

Publishers provided accepted manuscripts on a daily basis, with content volumes ranging from a few manuscripts to thousands in a given day. After processing, a centralised embargo management system within the PEER Depot released content to repositories on a daily basis in accordance with publication dates and embargo periods set for each participating journal.

Since PEER is a practical experiment, many of the solutions identified for the project can potentially be applied in ‘real life’ scenarios. PEER has developed, adapted and implemented a range of tools and technologies, many of which have potential applications outside of the finite duration of the PEER project including:

- Implementation of the SWORD protocol to allow application-level deposit of material into repositories
- Automated metadata extraction from manuscript PDFs (GROBID)
- Metadata mapping of different metadata schemas (NLM2.0, NLM 3 and proprietary formats)
- Establishment of 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)
- The creation of viable workflow models for content submission, filtering, processing and repository ingest
- An central embargo management facility to correctly manage the different embargo periods assigned to each journal
- Additional filtering of subject based content only for ingest by a subject repository (in the context of PEER this has been applied to social sciences content for SSOAR)

- Author deposit interface
- Online helpdesk with ticketing system where queries can be submitted and will be directed to the appropriate PEER project participant for a prompt response
- The ability to match author manuscripts with publisher provided metadata
- The development of a bug tracking & reporting workflow

These technological developments provide valuable practical outcomes from PEER.

#### References:

[1] Selection process-PEER participating journals;

<http://www.peerproject.eu/about/participating-journals/>

[2] *D3.1 Guidelines for publishers and repository managers on deposit, assisted deposit and self-archiving* <http://www.peerproject.eu/reports/>

[3] TEI and Scholarly publishing -experience from the PEER project, Laurent Romary, INRIA & HUB-ISDL

[http://dho.ie/sites/default/files/events/teipublishing/TEIScholPub\\_Dublin.pdf](http://dho.ie/sites/default/files/events/teipublishing/TEIScholPub_Dublin.pdf)

[4] Schematic of PEER Depot Workflow after *PEER Annual Report D9.8* (2010)

[http://www.peerproject.eu/fileadmin/media/reports/D9\\_8\\_annual\\_public\\_report\\_20100930.pdf](http://www.peerproject.eu/fileadmin/media/reports/D9_8_annual_public_report_20100930.pdf)

[5] Bretel F., Lopez P., Medves M, Monteil A., Romary L., INRIA & HUB IDSL  
*Back to meaning – information structuring in the PEER project* Author manuscript, published in "TEI Conference (2010)"

[http://www.peerproject.eu/fileadmin/media/ppt\\_about\\_peer/PEERBreakingNews.pdf](http://www.peerproject.eu/fileadmin/media/ppt_about_peer/PEERBreakingNews.pdf)

[6] Lopez. P. GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications. In *Proceedings of ECDL 2009*, 13th European Conference on Digital Library, Corfu, Greece.

### 3. Points of view: Publishers

#### BMJ Journals

BMJ Journals, generally, does not require authors to transfer copyright and permits public dissemination of accepted manuscripts. Authors are advised that they may post the accepted manuscript (but not the final published version of the Contribution unless the article has been 'Unlocked'), and the abstract of the final published article on a personal website and also on the website of any non-commercial employer. More generally, open access archiving of accepted manuscripts is permitted with an embargo of six months – via personal websites, institutional repositories and subject-based repositories.

Unlocked is a service that allows authors to make their articles freely available online, immediately on publication, for a fee. Unlocked is available to any author publishing original research in a BMJ Journal. On acceptance, authors will be asked whether they wish to pay to unlock their paper. Authors access to and use of Unlocked articles are covered by the terms and conditions of the exclusive licence agreement, which includes the terms and conditions of the Creative Commons Attribution Non Commercial 2.0 licence and specifically prohibits commercial use of these articles.

Open Access is not a big issue for BMJ Journals. Both green and gold open access is seen as compatible with the business model and complementary to the overall mission. Nevertheless, participation in the PEER Project was understood as a chance to observe change in scholarly communication and gauge demand for open access.

Transfer to the PEER Depot proved to be a fairly simple affair. BMJ Journals has experience with exporting accepted manuscripts via ftp, thus the PEER Depot was simply another destination. The manuscript is transferred upon acceptance and in a second transfer the DOI and date of publication are added. The only difficulty for BMJ Journals was the filtering of manuscripts according to EU authorship – but this task is performed by the PEER Depot.

The continuous transfer of accepted manuscripts has been smooth, and when trouble shooting was required, e.g. metadata was not validating, the PEER Depot and Project Management proved to be technically competent and prompt, so that any issue was resolved before becoming a problem.

Since January 2000 the BMJ Group and co-owners or contracting owning societies (where published by the BMJ Group on their behalf) have not asked authors of journal articles to assign us their copyright. Authors (or their employers) retain their copyright in the article; all we require is an exclusive licence (except for government employees who cannot grant this, thus non-exclusive) that allows us to publish the article in your chosen journal (including any derivative products) and any other BMJ

Group products, and allows us to sub-licence such rights and exploit all subsidiary rights.

We ask the corresponding author to grant this exclusive licence (or non exclusive for government employees) on behalf of all authors by reading our licence and inserting a relevant statement in the manuscript on submission.

Publishers are committed to the widest possible dissemination of and access to the content they publish. BMJ Journals supports any and all sustainable models of access that ensure the integrity and permanence of the scholarly record. Such options include 'gold' open access, whereby publication is funded by an article publishing charge paid by the author or another sponsor, a subscription-based journal, or any one of a number of hybrid publishing options. Most publishers now offer open access options and publish open access journals, and work closely with funders, institutions and governments to facilitate these developments. Gold open access provides one approach toward our shared goal of expanding access to peer-reviewed scientific works and maximizing the value and re-use of the results of scientific research.

BMJ Journals believes that authors should be able to publish in the journal of their choice, where publication will have the greatest potential to advance their field. Institutions and funders have a key role to play in ensuring that public access policies allow for funding of peer reviewed publication and publishing services in whatever journal that an author chooses.

BMJ Journals	
Total No of Journals	42
Open Access Journals	3
Open Access publishing option	37
Green Open Access option	0

BMJ Journals @ PEER	
No of participating journals	6
No of control journals	0

## EDP Sciences

EDP Sciences is one of the open access pioneers, with open access journals published since 2002. Currently, EDP Sciences publishes open access journals (based on article processing charges) as well as society-sponsored journals to which online access is free. Moreover, some journals have a 'moving wall': subscription-based content becomes accessible freely online after a specified period of time, depending on the field. For example, content in biology becomes freely accessible after one or two years, two years in astronomy and five years in mathematics.

EDP Sciences usually permits self-archiving by authors. EDP Sciences recommends that authors use the version of record (publisher's pdf) when posting to a personal website or a repository. This ensures that users download the correct version with the relevant journal details. EDP Sciences is rated a 'green publisher' by the SHERPA/RoMEO site. Currently, EDP Sciences does not impose any embargo.

Participation in the PEER Project was motivated by a desire to understand better:

- The motivation of authors and institutions when self-archiving;
- If open access repositories impact on subscription-based journals.

The transfer of manuscripts to the PEER Depot proved to be easy. For the participating journals, the Stage 2 manuscripts were extracted from the Manuscript Management System using a simple and logical procedure, similar to the ones used by EDP Sciences to send content (metadata & full-text) to different databases (e.g. Medline, ADS, CrossRef). The main difficulty for a medium-size publisher is the limited availability of resources and people, particularly when participating in a large international project.

The PEER Project has raised two issues for EDP Sciences. Firstly, if Green open access archiving becomes more prevalent, then publishers would be best placed to organize this deposit. It would be a service to the scholarly community, though the publisher must be able to recover any investment. EDP Sciences hitherto has permitted self-archiving because the manuscripts re-appear in dispersed locations. However, if partial and complete journals become available in disciplinary repositories, embargo periods should be set by domain and respected.

Secondly, the need for parallel archiving, or its value, is not necessarily clear. Publishers, individually and jointly, are investing in archiving, including fail-safe mechanisms that guarantee access in perpetuity should a journal no longer be published. Hence, it is unclear what the benefit to the scholarly community might be, if further copies are archived (at expense), particularly if these copies are not the version of record, but only the accepted manuscript.

EDP Sciences – academic and society journals	
Total No of Journals	42
Open Access Journals	6
Open Access publishing option	13
Green Open Access option	all

T&F journals @ PEER	
No of participating journals	2
No of control journals	0

## Elsevier

One of Elsevier's primary missions is to work towards providing universal access to high-quality scientific information in sustainable ways. We therefore felt it very important to participate in the PEER Project as evidence-based policy making is invaluable for all stakeholders.

Elsevier is committed to providing the broadest possible access to its publications, whilst at the same time upholding the highest level of quality. This means significant, continued investment in the publication system. As this system develops new business models will emerge, and we are very happy to use any sustainable model. Open access business models have a role to play as part of a diverse landscape that also includes other business models including the proven licensing and subscription model.

Elsevier has engaged with open access for a number of years. We have a long-standing record of working cooperatively and successfully with funding bodies to provide open access options (e.g. the Wellcome Trust, and Medical Research Council UK, FWF in Austria and Telethon in Italy), and we publish 13 open access titles, 1200 hybrid journals, and over 40 journals with open archives. Our open access initiatives are outlined online at: [http://www.elsevier.com/wps/find/intro.cws\\_home/open\\_access](http://www.elsevier.com/wps/find/intro.cws_home/open_access).

Elsevier's main motivation for participating in the PEER Project was to further explore the costs and sustainability of green open access models, which focus on manuscript posting but with no revenue stream. Elsevier is classified as a green publisher and is supportive of individual and voluntary deposit of preprints and accepted author manuscripts by authors from the time of publication. Where an employer or funder has a mandate in place we ask that the institution enter into a systematic posting agreement with us and to use an embargo period after which the content can be made publicly accessible. Elsevier has developed a number of these "green" manuscript posting arrangements with institutions (e.g. RKI, NRC-Canada, Czech Academy of Sciences).

Elsevier has specific concerns about some approaches to systematic green open access – particularly approaches which rely on short and/or inflexible embargo periods without some form of remuneration to the publisher to offset publication costs. In some cases we already see some reduction of usage (by subscribers) and transactional sales (for non subscribers) for articles on our publishing platform and we do not believe these models are sustainable.

In the PEER Project, Elsevier has provided much content directly to the PEER Depot. Most aspects worked reasonably well but there were some aspect that required more work than was anticipated initially. All issues were resolved but it may be helpful to list the main issues in setting up content transfer to the depot:

- (1) The interface between Elsevier's workflow and the PEER Depot  
All metadata required by the PEER depot, had to be converted by an external supplier into a suitable format for the depot. This supplier was a point of

disconnection between our workflow and the depot. Managing this added a further level of complexity to the process.

#### (2) Identification of manuscripts and the metadata workflow

During the lifetime of the project a number of articles came to light for which the depot had received a PDF from the author but with no matched metadata. Without metadata these would remain as “unmatched manuscripts”. It proved very hard to get to the root cause of this, but eventually we settled for a pragmatic approach in which the required metadata was redelivered to the depot, after conversion via a third-party supplier.

#### (3) Embargo management

Embargo times were managed from within the PEER depot, based on a simple spreadsheet we have provided. For some older material this required manual work-arounds and the use of print publication dates, which may have delayed some articles. However, we delivered a large number of older articles as backfiles, which helped alleviate this.

#### (4) Logfile provision

Providing logfiles for 106 journals (out of over 1800 on our site) required setting up a separate project to extract and provide usable log data for the journals in the PEER project. Providing raw logs was not practical or possible as this would have included the logs for all 1800+ journals across all years measured.

Elsevier – academic and society journals	1882
Total No of Journals	1882
Open Access Journals	12
Open Access publishing option	More than 1200
Open Archives	40+ titles
Green Open Access option	All journals with the exception of a few titles

Elsevier journals @ PEER	106
No of participating journals	53
No of control journals	53

## IOP Publishing

IOP Publishing has been a gold open access publisher for more than ten years, since before the phrase came into common usage. In 1998, in partnership with the Deutsche Physikalische Gesellschaft, it launched as an experiment the pure open access journal *New Journal of Physics*. It has since launched six more pure open access journals, including *Environmental Research Letters*, *Science and Technology of Advanced Materials* and three conference series.

It has also now launched a 'hybrid' gold open access option on 23 of its owned subscription journals, along with three journals published in partnership with other organisations. It is talking to other partners about offering the hybrid option on other co-published journals too. It is making the hybrid option available alongside the pure open access option so that researchers who wish to publish their research in IOPP journals, and who wish to do so on an open access basis, or are required to do so by their funding bodies, can do so in their journal of choice. IOPP will take revenues from publication fees fully into account when setting subscription prices for hybrid journals.

Authors have been supported in posting their accepted manuscripts to personal websites. IOPP also supports authors who are required by their funding agencies to make their research papers freely available via an institutional or subject repository. Authors may post their accepted manuscript in an institutional or subject repository after an embargo period of 12-24 months following publication, depending on the journal. IOPP believes that embargoes are required before such posting in subject or institutional repositories because this green form of open access makes no contribution to the costs of publication, including the management of peer review, and systematic posting could undermine the services it provides in scholarly communications in physics. It believes that gold open access is a much better option, as it provides immediate access to the final published version of the article.

We regard arXiv as different to most subject repositories. It began as and remains predominantly a preprint server and there are some basic workflow aspects to it that are central to some of our research communities. We are therefore happy to work with it and we enable authors to input their arXiv number on submission.

It should be noted that IOPP also makes most of the articles that it publishes freely available for the first thirty days after publication. This policy remains under regular review as it may become unsustainable if funding agencies require accepted manuscripts to be made freely available after short embargo periods, thus reducing the window in which IOPP can earn back its investment in its publication services.

IOPP participates in PEER because it aims to provide an evidence base for the technical and cultural issues involved in posting manuscripts to a green open access repository and the impacts on usage and traditional business models of such posting on a large scale.

In its interaction with PEER, IOPP encountered several challenges, some to do with the set-up of a central depot, some more generally with capturing the accepted manuscript for deposit. The main issues were:

- It took the PEER Depot quite a while to develop a stable transfer routine between IOPP and the PEER Depot, and in the end there was still no match between the metadata fields required by the depot and those used by the publisher, so that the metadata are extracted from the file transferred by the publisher;
- Some of the metadata required by the depot are not available in the IOPP manuscript management system, but only added in the publication system, so that manuscripts can be sent to the depot not upon acceptance, but only after publication;
- Locating and extracting manuscripts by EU authors only is a challenge in itself, and is not possible consistently unless the corresponding author has an address in a EU country.

Overall, the PEER Project took up more time for staff in IT, management and administration than IOPP anticipated.

It remains problematic for IOPP that green open access is heavily reliant on the work that publishers carry out in organising the review, editing and publishing of manuscripts. Experience of the PEER Project to date appears to show that any large-scale solution would require the active involvement of publishers, not least because authors do not appear to deposit in repositories in any significant numbers even when encouraged to do so.

IOPP – academic and society journals	
Total No of Journals	65
Pure Open Access Journals	7
Hybrid Open Access Journals	27
Green Open Access option	After embargo period

IOPP journals @ PEER	
No of participating journals	6
No of control journals	6

## Nature Publishing Group

Nature Publishing Group (NPG) does not require authors to transfer copyright and it encourages self-archiving. NPG also offers a free manuscript deposition service to authors subject to a funders open access mandate (e.g. PMC and UKPMC). The NPG author licence policy states that

Authors grant NPG an exclusive licence to publish, in return for which they can reuse their papers in their future printed work without first requiring permission from the publisher of the journal...

When a manuscript is accepted for publication in an NPG journal, authors are encouraged to submit the author's version of the accepted paper (the unedited manuscript) to PubMedCentral or other appropriate funding body's archive, for public release six months after publication. In addition, authors are encouraged to archive this version of the manuscript in their institution's repositories and, if they wish, on their personal websites, also six months after the original publication.<sup>2</sup>

NPG is also an open access publisher. Notably, Nature Communications (launched April 2010, APC €3570) has seen an uptake of forty percent open access. Scientific Reports (June 2011) is fully open access.<sup>3</sup> In addition, NPG offers open access options on 37 of its academic and society journals and has an additional six open access journals. Open access articles are published under a choice of Creative Commons licenses.

NPG was already exporting to PMC and UKPMC and was thus interested to use the same procedure for PEER. The green version of the article was exported from the manuscript tracking system. However, for PEER some functionality had to be added, namely by adding the DOI and the publication date for any submission to the PEER Depot. Hence, for NPG the initial setup was quite straightforward and the transfer of manuscripts to the PEER Depot has worked smoothly.

In the interaction with the PEER Depot and the repositories, two issues were notable for NPG, namely that

- It took a while for the repositories to appreciate that each publisher had a different system and that some faced considerable technical difficulties in capturing and transferring the accepted manuscript;
- Many repositories had not developed any functionality to handle embargoes.

Looking ahead, NPG expects open access publishing to grow and become increasingly significant – more so than green open access. NPG is observing a shift in authors' attitudes, whereby open access publishing is increasingly accepted as a fast and efficient form of scientific communication. With regard to Green Open Access, the level of support that PEER provided, and particularly the PEER Depot, has been the best that NPG has experienced. Going forward, for any Green Open Access scenario, it would be imperative that the same level of support is maintained.

---

<sup>2</sup> <http://www.nature.com/authors/policies/license.html>

<sup>3</sup> [http://www.nature.com/press\\_releases/statement.html](http://www.nature.com/press_releases/statement.html)

NPG – academic and society journals	
Total No of Journals	61
Open Access Journals	11
Open Access publishing option	46
Green Open Access option	all

NPG journals @ PEER	
No of participating journals	16
No of control journals	16

## **Oxford University Press**

OUP has been a leader and an innovator in open access publishing, both in terms of fully open access journals and with a hybrid open access program, Oxford Open. Experience and findings have been shared with the scholarly community on a regular basis through talks at conferences and articles in industry publications. Open access fits with OUP's central mission of maximising the dissemination of high quality scholarly information and exploring new publishing models, whilst at all times ensuring the long term interests of our authors and society publishing partners are kept firmly in mind.

OUP does not have a particular stance towards open access. OUP experiments responsibly and responds to the market. Where clients prefer OA publishing options, OUP provides them. OUP is aware that there is currently no one size fits all open access publishing or business model, and that what works in medical publishing may, for example, not be applicable to humanities publishing. OUP runs various OA models, from sponsorship (e.g. Journal of Legal Analysis) to article-processing charges (e.g. Database, Nucleic Acids Research) to author self-archiving.

OUP is very conscious of the ethical aspects of open access – we ensure that all editorial decisions on articles in hybrid journals are taken before the open access option is offered to authors, and we actively reduce online subscription prices in line with open access uptake.

OUP was happy to participate in the PEER as the motivations and aspirations of the project were in line with our overarching strategy regarding OA – experimentation, innovation and sharing findings. OUP did not have any pre-conceived notions on what the project would bring but was a willing participant.

Participation in the PEER Project and the transfer of accepted manuscripts to the PEER Depot was straightforward. Once the routine had been developed, the transfer was efficient – just as OUP has found with transfers to PMC. However, much time elapsed until the routine was in place because the PEER Project changed and adjusted guidelines several times.

OUP had agreed a list of six journals for participation in the PEER Project – in consultation with the internal OUP publishers and external journal editors. Scholars publishing in these journals needed to be informed. The wording of the letters to authors and on the Scholar One sites (peer review system for submitted manuscripts) was agreed. A new ftp delivery address was added to the six sites, the agreed text inserted, and instructions placed for authors. For each journal a monthly report was sent to PEER, listing which manuscripts PEER should have received. This task was more complex as information from the production tracking system was needed. The solution was a bespoke report for PEER.

This last task, reporting from production, was the only thing OUP had to build from scratch. No other challenge was encountered – with the caveat that only six journals participated in PEER.

Overall, the experience with the PEER Project was very much as OUP expected: Fairly difficult to get a consensus amongst publishers, and unlikely to get many files

from authors without mandates. The behavioural analysis conducted by PEER was really telling in confirming some latent thoughts on author behaviour, such as that

- a) The increase in the number of authors reporting placing their article in an Open Access repository was negligible;
- b) Readers had concerned about the validity of the version of the article they were viewing in a repository.

Open questions that remain are: Was the level of author education about the purpose of the exercise right? If usage levels at repositories average five percent of those at publishers, does this level of alternate or additional usage justify the investment in repositories?

Overall, OUP has the impression that Green open access is much more likely to get a good level of compliance if the publisher deposits: It makes sense and should be more efficient. One would think that would be what the author would want to - but on ethical grounds OUP believes that authors should deposit their manuscript in a repository of their choice. More specifically, transfer of content is easier if the journals are on ScholarOne. If had not been possible to extract manuscripts from the peer review system, it would have been significantly more laborious.

On the other hand, the PEER findings, particularly the low level of author compliance when it's optional (and there's no specific benefit for them), are very instructive. That raises questions about how many authors actually want Green OA and how many are doing it just because they are compelled to.

OUP – academic and society journals	
Total No of Journals	283
Open Access Journals	9
Open Access publishing option	110
Green Open Access option	281

OUP journals @ PEER	
No of participating journals	6
No of control journals	6

## Portland Press

Portland Press is the publishing subsidiary of the Biochemical Society (which is devoted to the cellular and molecular life sciences). It publishes the journals of the Biochemical Society as well as journals for other organizations and provides a range of publishing services for its clients. Portland Press has just converted the Biochemical Society's *Bioscience Reports* to open access and also in 2009 launched the open access *ASN NEURO* for the American Society of Neurochemistry. Earlier, in the late 1990s, Portland Press experimented with a moving wall, whereby content of the preceding calendar year was made freely available.

In 2006, the *Biochemical Journal* back archive (1906-1996) was digitized and deposited in PubMed Central as part of the Wellcome/JISC/NLM digitization programme. That same year, Portland Press experimented with a rolling 6-month release on the journal's website and to PubMed Central for the *Biochemical Journal*. However, when Opt2Pay (see below) was introduced in 2008, the journal returned to a delayed-access model on a rolling 12-month basis and only articles prepaid for by authors were deposited in PubMed Central (with posting to PubMed Central mirror sites), because of concerns about loss of usage from the journal site to PubMed Central.

For authors publishing under an open access mandate, the Opt2Pay model allows payment of an article-processing charge that makes the version of record freely available immediately under a Creative Commons licence in Portland Press journals. Portland Press also permits self-archiving of the author's final accepted manuscript in institutional repositories 6 months after publication. The author's final manuscript can be deposited in PubMed Central upon publication with an embargo on public release for 12 months. The time at which content in Portland Press's journals becomes freely available varies: for example for the *Biochemical Journal* it is 12 months after publication.

Although Portland Press is willing to work with the research community in increasing accessibility to published research, including through allowing green open access, it has some concerns about the possible impact on subscriptions of free availability of published papers (whether the author's final accepted version or the version of record). Libraries are now using usage data when making purchasing and renewal decisions, which means that downloads at repositories implies lost usage. It is increasingly clear that large disciplinary repositories do have the potential to divert usage.

Portland Press supported the PEER Project as it provided an opportunity to collect data about authors' attitudes and user behaviour. The transfer of manuscripts to the PEER Depot proved simple because the online peer review system (developed in-house) provides an infrastructure whereby the author's final manuscript is posted online immediately after acceptance as a PDF. It was therefore straightforward to transfer this PDF and the accompanying metadata for participating journals.

Overall, the PEER Project has enabled Portland Press to engage with the changing open access landscape. Noteworthy has been a general (though still slow) process of education among scholars and stakeholders, leading to a more realistic appraisal

of the costs and opportunities associated with open access. For Portland Press, three points have emerged:

- If authors prefer (or are obliged) to make their written work available through open access, then (gold) open access publishing is the more straightforward and lower-risk route;
- For any Green open access scenario, publisher-assisted deposit would be preferable to ensure version control and an understanding of where papers are deposited and the extent of the activity;
- If publishers are requested to cooperate with repositories (i.e. via open access repositories), then an accepted way of adding up usage at different sites should be developed, for example by repositories providing COUNTER compliant usage statistics.

Portland Press journals	
Total No of Journals	8
Open Access Journals	2
Open Access publishing option	Opt2Pay in subscription journals
Green Open Access option	Portland Press also permits self-archiving of the author's final accepted manuscript in institutional repositories 6 months after publication. The author's final manuscript can be deposited in PubMed Central upon publication with an embargo on public release for 12 months.

Portland Press journals @ PEER	
No of participating journals	3
No of control journals	0

## Springer

Springer Science+Business Media has the largest portfolio of open access journals worldwide. With well over 300 titles, the brands SpringerOpen and BioMed Central offer a wide and quickly growing range of fully open access journals.

In addition, Springer provides an open access publishing option – Springer Open Choice – in almost all its subscription journals. This enables authors to choose the publishing model after their article has been through the peer review process. All open access content at Springer is published under the Creative Commons Attribution (CC-BY) license, which permits commercial and non-commercial re-use of an open access article as long as the author is attributed.

Springer fully supports – and significantly invests in – open access as a business model, and welcomes any opportunity to develop and grow this model in partnership with researchers, institutions, societies and foundations.

“Green” open access archiving does not cover the costs associated with formal publication, and in our view poses risks in terms of the sustainability of scholarly communications. Nevertheless, Springer always tries to assist authors in meeting publishing requirements they may face, and therefore allows them to post the accepted manuscript of their articles on their personal website or institutional repository. Authors may also deposit this version on their funder’s or funder’s designated repository at the funder’s request or as a result of a legal obligation, provided it is not made publicly available until 12 months after official publication.

Springer was pleased to participate in PEER in order to gain experience and evidence for the technical issues and attitudes involved in author-posting to repositories, as well as for costs and benefits. The interaction with the PEER Depot was fairly straightforward and worked well, but we think it useful to mention a few practical issues. For example, “accepted manuscripts” are a version of the article that we do not usually record, therefore the production system had to be tweaked to accommodate the project. This sounds trivial, but requires attention and resources. As another example, it seems that many repositories do not have processes to manage embargoes.

These examples illustrate our main conclusions from PEER: The technical and administrative challenges involved in depositing manuscripts in repositories are far greater than is usually presumed. The PEER Project took up more time in IT and management than even we had expected. Considering that the PEER Depot provided much greater support than repositories generally do, we believe that technical, administrative and standardisation issues are a real challenge for Green OA initiatives. If authors were left to deal with deposit by themselves, these issues would remain difficult to resolve.

Looking ahead, Springer expects Gold open access publishing to increase significantly. The most efficient way to approach repository scenarios might be to integrate publisher-assisted article deposit into open access publishing options. For example, Springer deposits final open access articles into repositories of funding institutions via SWORD protocol.

Springer journals incl. BioMed Central and Springer Healthcare	
Total No of Journals	3.192
Open Access Journals	363 (fully OA journals published by SpringerOpen, BioMed Central, Chemistry Central, Springer Healthcare)
Open Access publishing option (Springer Open Choice)	1.435
Green Open Access option	2.829 (all subscription and hybrid journals)

Springer journals @ PEER	
No of participating journals	27
No of control journals	27

## Taylor & Francis

Since 2006, Taylor & Francis has had an open access policy, which it continues to review. Currently, the self-archiving of accepted manuscripts is permitted, subject to an embargo of twelve months for STM and the behavioural sciences (e.g. psychology) and eighteen months for the social sciences and humanities. The embargo period for the behavioural sciences was reduced from 18 to 12 months in 2009. Taylor & Francis continues to experiment with green open access postings via our author rights policies. Currently, 550 journals offer an open access option (APC €1900).

The transfer of accepted manuscripts from the editorial management system (ScholarOne) to the PEER Depot has proved to be relatively straightforward but fairly time-consuming. Also, the modus operandi of the PEER Project (direct publisher submission) has proven less labour intensive than the process for PMC - with more manual technical intervention and more communication with authors. However, it should be noted that the costs for the setup were quite high to T&F and the in-house coordination required non-trivial. Some of the major challenges were:

- Between the PEER Depot and T&F the parameters (metadata) for the transfer of content had to be agreed, and this required lengthy negotiations;
- Subsequently, to meet the requirements of the PEER Depot, complex and time consuming in-house coordination was required to tweak the system and export the accepted manuscript;
- A second export to the PEER Depot had to be setup to pass on the DOI and date of publication of each article.

Once the setup is complete, the export of accepted manuscripts is relatively routine, although the transfer of accepted manuscripts is straightforward only when going forward, not for any back content.

Key to open access, for T&F, is that any route is feasible and viable. Currently, open access scenarios seem fluid, and further investigation of potential costs and benefits required. The PEER Project has been working well. Taylor & Francis believe that publisher and publisher-assisted deposit may be ways forward, provided that the embargo is respected. However, for this to be feasible, the requirements for the transfer of content would need to be standardized, preferably across all repositories.

T&F – academic and society journals	
Total No of Journals	1624
Open Access Journals	10
Open Access publishing option	680
Green Open Access option	1624

T&F journals @ PEER	
No of participating journals	39
No of control journals	40

## Wiley

Wiley has been sceptical of efforts to mandate open access, particularly by asking researchers to deposit the accepted manuscript (green open access). Though some journals permit self-archiving, Wiley generally does not support green open access – because it is seen as dependent on the services and revenues of the subscription-based publishing model, from which it takes but does not contribute; it leads to “version confusion”; and it diverts usage from the publisher’s site. Wiley makes an exception for the NIH (National Institutes of Health in the USA) mandate because it is the only one in the world backed by statute. As most other publishers offer to deposit to PMC on behalf of the authors, Wiley currently does so too. The embargo for accepted manuscripts is twelve months after publication. Wiley has been experimenting with hybrid open access publishing, and as of 2011 is launching a fleet of “Gold” open access journals.

As Wiley is critical of open access mandates, the main motivation for participating in the PEER Project is to explore the difficulties that arise in mandating the deposit of accepted manuscripts by researchers. Moreover, the PEER Project simulates a possible mandate at the European level, and allows for some insight into issues of implementation as well as possible consequences. It is not Wiley’s intention to make the process artificially easy for the repositories since that would not emulate a real-life situation.

Wiley has noted with interest that the PEER Project had to implement direct publisher deposit to make the project viable because author self-archiving, even when permitted and encouraged by publishers in writing, was expected to be low – and did indeed prove to be minimal. Moreover, publishers had to deposit into a central depot, which filtered and processed the manuscripts, because institutional repositories did not have the capabilities.

As regards implementation, Wiley has been able to implement an automatic routine that captures the manuscript upon acceptance and transfers it to the PEER Depot, with the subsequent addition of the DOI and date of publication. Principally the setup is straightforward, but it does result in significant coordination and implementation costs (order of magnitude: six-figure-sum).

From Wiley’s perspective, within the PEER Project, publishers have had to offer too much support to institutional repositories – through the transfer of metadata and content to a depot – confirming that green open access needs publishers in order to work efficiently. First of all, the methodology for capturing the ‘final peer-reviewed manuscript’ depends on publishers and the peer review services they organize. Secondly, the publishers must provide an array of metadata for the management of the manuscript and the embargo period. Ultimately, the publishers get drawn into depositing manuscripts because the large majority of authors will not self-archive. Hence, green open access is not considered a viable way forward unless there is some licensing scheme that can be agreed between publishers and repositories.

Wiley – academic and society journals	
Total No of Journals	1500
Open Access Journals	14
Open Access publishing option	1374
Green Open Access option	Varies according to journal ownership and subject discipline behaviour (eg attitudes to the circulation of un-peer-reviewed or unpublished material). Wiley's default position is that the submitted version can be deposited, usually after publication; accepted version and published version can't be deposited

Wiley journals @ PEER	
No of participating journals	58
No of control journals	60

## **4. PEER from the perspective of the repositories: Experience, outcome and conclusions for open access policy**

From the perspective of the participating open access repositories (OAR), the outcome of the PEER experiment can be examined on two levels. First, with regard to the quantity and quality of scientific articles processed for open access repositories (DEPOSIT MODEL). Second, by evaluating the implementation of an efficient (i.e., automated) transfer process for large scale deposit of stage-two manuscripts and metadata from a number of participating publishers to a number of distributed open access repositories (TRANSFER MODEL). For the repositories, the PEER Observatory is a comprehensive test-bed for the development of processes, implementation of solutions and for cooperation between the participating actors. We asked the repositories about their experiences with the deposit and transfer process and the technical solutions in the PEER Observatory, their appraisals of each individual solution and which aspects they think will become a part of future open access models. OARs were interviewed and surveyed between May 2011 and March 2012. This report outlines the experiences and conclusions of the repositories.

### Structure of the report

1. Evaluation of the PEER deposit model
2. Evaluation of the PEER transfer process and the technical solutions
3. Outcome for the participating repositories and conclusions for a future OA strategy

#### 4.1 Evaluation of the PEER deposit model

The deposit model of PEER is characterized by two principles which are not related necessarily: the systematic, automated provision of a large number of manuscripts by publishers on the one hand, and the deposit of stage two manuscript versions on the other.

##### *Systematic, large-scale deposit by publishers*

From the perspective of the participating repositories, large-scale deposit by publishers is a very convenient model for the implementation of Green open access. In particular, the provision of metadata was rated very positively. The following table shows details of these assessments. The column ‘Mean’ indicates the mean rating by five repositories on a scale from 1 ("very good") to 5 ("very poor").<sup>4</sup>

<i>T 1: How would you rate the deposit model of PEER?</i>	<i>Mean</i>
The provision of metadata in PEER	1.4
The deposit of authors' manuscripts by the publishers	1.6
The systematic deposit of large numbers of manuscripts	1.6
The automated transfer and checking of large numbers of manuscripts in PEER	1.6
The number of manuscripts that PEER provides for your repository	1.8

Positive outcomes from the perspective of OARs are for example: *"the systematic coverage of published articles"*, *"to get scientifically valuable, peer-reviewed content"*, *"administration removed from authors; no burden for authors"*.

The repositories were more critical of the outcome of PEER with regard to the number of manuscripts actually available to the repositories. The number of articles that were disseminated in the end from the PEER Depot to the OARs is restricted by: *"Large embargo times"*, *"publishers problems in providing stage-2 manuscripts"*, *"uneven quality of data"*.

All repositories agree that systematic deposit by publishers should be part of future green open access policies (mean: 1.4). This is because, in spite of the limitations, PEER considerably expands the number of journal articles available to the OARs through open access: by mid-April 2012 the PEER OARs had received a total of 18,556 articles from the PEER depot. The subject-based SSOAR received 3,100 articles from PEER.

Other advantages from the perspective of the repositories include the tested and standardized metadata for all documents and the automated transfer of documents to the distributed OARs.

##### *Self-archiving by authors*

<sup>4</sup> Given the small sample, the mean is a qualitative indicator.

Another reason that publisher deposit was rated positively by the repositories is that the rate of self-archiving by authors was very low. One could consider whether the "invitation" to self-archive and self-archiving process itself could be improved, but most of the repositories prefer solutions where the publisher deposits the article on behalf of the author, if the author agrees or is mandated to provide open access to the article.

### *Stage two manuscripts*

A stage two manuscript is a final, peer-reviewed author's version, but the document is transferred from the publisher's internal manuscript or production system. There could be different versions available in the last stages of the publisher's editorial process. Clearly, Stage II is not an exactly defined version and, what is even more important, it is originally not intended for public distribution.

Stage II manuscripts can be very heterogeneous; the version of the PEER manuscripts differs according to publisher, journal and time period. The critical point here is that some of the versions provided by some publishers contain information about the status of the editorial process, such as stamps indicating the internal processing status ("for peer review only," "confidential," etc.). Such versions can be confusing for readers, in particular if they want to cite the article (see repository user survey). The authors, too, probably do not appreciate having these articles made publicly accessible – some repositories have reported receiving complaints about this from authors. Some of the manuscripts included comments, corrections or annotations made by authors, referees and editors and, consequently, had to be removed from the repositories.

Provision of stage two manuscripts was the aspect that received the skepticism from the repositories. It received the lowest rating of all aspects in the survey, as indicated by the mean of 5 PEER repositories on a scale from 1 ("very useful") to 5 ("not useful at all").

<i>T 2: How would you rate:</i>	<i>Mean</i>
the provision of stage-two manuscripts from publishers for your repository	2.0
the quality of the stage-two manuscripts in PEER is	2.4

Note, however, that Stage II manuscripts are not generally criticized. Stage II manuscripts were rated "very useful" by three repositories. The subject repository, which exercises systematic quality control, rated Stage II manuscripts as not useful. Negative experiences were reported as follows:

*"Sometimes reviewer comments and uncompleted deposits drive to withdrawal of articles - for fully automated procedure this is sometimes a problem, as there are no resources to review the completeness and clearance from reviewer comments."*

*"In PEER, some manuscripts provided by publishers contain confidential comments from the peer-review process. Publishers should check what they deposit."*

*"Authors don't like to share their stage two manuscripts."*

The repositories were more reserved to whether Stage II manuscripts should be part of a future green open access policy. Only one repository chose the answer, "yes, absolutely," and that was because they see the deposit of Stage II documents as an "important step in the open access publishing process." In principle, Stage II manuscripts meet the requirements of the green open access policy. But after their experience with PEER, the majority of the repositories find Stage II manuscripts to be only conditionally suitable – e.g., "*when no stage-3 version is available (preference for publishers' PDFs)*".

The problematic point in the PEER deposit model is the combination of heterogeneous quality of publishers' Stage II documents on the one hand and the automated transfer process without systematic quality checks by repositories on the other.

Typically, quality checks were made by the OAR's when (green OA) manuscripts are delivered by authors (self-archiving). Those quality checks are done manually, therefore they are laborious and time-consuming for repositories. In PEER, quality control procedures for manuscripts are difficult to apply due to the large number of documents. Quality control checks are not provided for in PEER and thus cannot be done with the resources (personnel) that PEER has at its disposal. Only SSOAR systematically checks every manuscript before it is added to the repository, but as a subject repository these are "only" 3,100 manuscripts in contrast to roughly 19,000 in the other repositories. To date, flawed manuscripts have to be removed manually from the PEER repositories. Since this action has been repeatedly found to be necessary, several repositories have suggested implementing an update or remove function in SWORD.

## **4.2 Evaluation of the transfer process and the technical solutions implemented by the repositories**

One central goal of PEER from the perspective of the repositories was the development, implementation and intensive practical testing of an automated transfer process so that large quantities of documents and metadata provided by publishers, can be processed, and – ideally without further manual intervention – disseminated to a range of participating repositories. The PEER transfer process is supposed to implement efficient procedures, a suitable technical platform, standardized processes, (meta-)data structures and technical solutions. These could support the spreading of open access, within and beyond the PEER experiment.

The main elements of the PEER transfer model are the SWORD protocol, implemented as the standard transfer at all participating repositories, and the PEER depot as central technical platform, clearing house, processing unit and dark archive. This central and intermediary unit in PEER makes it possible to have the documents (plus metadata) from publishers continuously added to the distributed open access repositories, with no further manual interaction necessary.

In addition to implementing SWORD and TEI, the repositories accomplished other tasks as well, providing such services as a "help desk" for users, an "Author Manuscript Deposit at the Help Desk" interface, and tools for "PEER Depot Reporting" (MPDL), as well as sending

logfiles to the usage research team and looking into possibilities for long-term preservation of PEER content (INRIA / PEER depot).

The success of technical solution developed in PEER is manifest from the on-going transfer to the repositories, which is handled by the repositories with minimal time and effort (a couple of hours per week). All repositories gave very positive feedback concerning the transfer process, as seen from the following quotes taken from the interviews (final questionnaire) with the repositories:

*Processes technically went well, once journals were agreed on, delivery was constant and fluent, fulltexts came already with metadata; From the aspect of a repository, the technical solution is pretty fine, as we get manuscripts deposited by the SWORD interface; Automated ingest; embargo management done by PEER Depot; very smooth workflow; Centralized and well organized automatic depositing of large amount of data with pre-filtering; Large collections of manuscripts, established automated repository deposit; SWORD protocol is very effective; good observation of data flows; Depot as intermediate processing unit facilitates publisher deposit process; one central place to collect and distribute; Useful to have a unit what can handle such administrative jobs: to match metadata with the manuscripts, take care of embargo period, communicate with publishers, automatically deposit data into several repositories;*

The positive assessment applies in particular to the use of the SWORD protocol as the standard for transferring documents to repositories (see Table 3).

<i>T 3: Technical solutions</i>	<i>Mean</i>
the transfer process via SWORD protocol	1.2
the process of matching manuscripts with metadata by the PEER depot	1.4
the embargo management in PEER	1.4
the PEER Depot as clearing house, processing unit and dark archive	1.6
the metadata specifications (TEI) in PEER	1.8
the process of checking manuscripts by the PEER depot	2.6

Embargo management by the PEER depot was also rated very positively (1.4), as was the process of matching of documents with metadata. The PEER depot as central processing unit for the transfer process was positively assessed by all repositories as a highly efficient solution. Table 3 shows the individual ratings of the technical solutions.

The standard introduced for handling metadata, Text Encoding Initiative (TEI), is seen less positively; in particular, generating TEI data for the PEER depot from the metadata provided by the publishers seems to be difficult. The transformation of publisher metadata into a unique exchange format (TEI), and the maintenance of consistent transformation of publisher metadata to TEI (XML testing framework) was one of the main problems to be overcome by the PEER depot. The result was not optimal, neither from the point of view of the PEER Depot nor from that of the repositories (1.8).

The most important outcomes of the technical solutions in PEER from the point of view of repositories are:

- the SWORD protocol for exchange of documents and metadata
- the automated transfer/deposit of documents and metadata
- the implementation of a central processing unit (PEER depot)

All repositories stated that they will use the SWORD protocol in future. Also, the concept of the PEER Depot as a central processing unit and clearing-house should be applied in future open access scenarios. In this context, it was suggested that a central depot for open access repositories should be provided on a long-term basis: *"Repositories could 'subscribe' to the central depot, and then specify the subset of publisher deposits they would like to receive. Also, the option of defining selection criteria for accepting manuscripts in the repository could be of benefit for repositories."*

### 4.3 Outcome for the participating repositories and conclusions for a future OA strategy

From the perspective of the repositories participating in PEER, the outcomes can be considered on several levels. First of all, there is the outcome for each repository in terms of open access content. Second, there are outcomes for participating repositories in terms of the technical standards implemented, enhanced infrastructure and technological expertise. And third is the outcome for future open access policy.

We asked each repository to rate certain PEER outcomes (Table 4) on a scale from 1 to 5 (1 = "yes, absolutely"; 2= "yes"; 3= "to some extent"; 4= "probably not"; 5= "absolutely not"). Compared with the ratings of the technical solutions and the deposit model in general, the outcome for each repository is seen more critically, in particular regarding the restriction to corresponding EU authors and the scope of journals in PEER.

<i>T 4: The outcome of PEER for your repository</i>	<i>Mean</i>
Does the PEER content improve your repository?	2.0
Are the articles by EU authors valuable for your users?	2.2
Is the scope of journals and manuscripts in PEER valuable for your users	2.6

The most important outcome of the PEER project for the participating repositories is the implementation of SWORD, improved knowledge of handling of SWORD, the experience with the central processing unit (PEER depot) as a technical platform for automated transfer, an experience of automated deposition by publishers, automatic matching with DOI and the use of TEI and SWORD. And in this respect, as well, PEER is quite successful from the repositories' perspective—as Table 3 above shows. Repositories will use these technical

solutions as widely as possible. Moreover, the PEER experience fostered cooperation between repositories on a European level.

What have repositories learnt from PEER for future open access policy?

By all means the deposit by publishers should be part of future open access policies, as should SWORD (and TEI) as standards for transfer and exchange of OA documents and metadata (Table 5).

<i>T 5: Conclusions for future OA policy</i>	<i>Mean</i>
Should the deposit by publishers be part of future green open access policies?	1.4
Should stage-two manuscripts be part of a future green open access policy?	2.2
Should the SWORD protocol be part of a future green open access policy?	1.2
Should a central depot for the transfer process be applied to future green open access solutions?	1.8
Should the PEER model be part of future OA policies?	2.0

A central processing unit has proved to be effective but on a practical level, this functionality depends on the specific open access scenario. More controversial is the deposit of stage-two manuscripts – either they should be checked before transfer to the repositories or defined exactly before deposit or substituted by the final publisher version.

## 5. Author Deposit: Numbers and reflections\*

The PEER Project was based on making available - experimentally - a significant number of authors' final peer reviewed manuscripts at the European level: to observe process and effects. The build-up of the infrastructure was described earlier: principally, half of the eligible manuscripts were transferred directly from publishers, and for the other half the authors were invited to self-archive. However, author deposit rates were low; consistently low throughout the lifetime of the project. By February 2012, the PEER Depot counted only 170 deposits by authors (from >11000 invitations in 21 months from December 2009 to August 2011).

The main characteristics of the PEER author deposit infrastructure were:

- Publishers raise author awareness: Upon submission, authors received notification that the journal they were submitting to was participating in PEER and that if accepted, they may be invited to self archive their manuscript (most but not all publishers sent this message);
- Publishers invite authors: Upon acceptance in a participating journal, authors received an invitation to self-archive their manuscript;
- Single deposit interface with a help desk: Authors were provided with instructions for deposit into the PEER Depot and supported by a help desk;
- Centralized management: The PEER Depot processes the manuscript and handles the embargo.

The invitation to authors read as follows:

"This journal is participating in the PEER project, which aims to monitor the effects of systematic self-archiving (author deposit in repositories) over time.

<http://www.peerproject.eu/>

PEER is supported by the EC eContentplus programme,

[http://ec.europa.eu/information\\_society/activities/econtentplus/index\\_en.htm](http://ec.europa.eu/information_society/activities/econtentplus/index_en.htm).

As your manuscript has been accepted for publication by [Journal name], you may be eligible to participate in the PEER project. If you are based in the European Union, you are hereby invited to deposit your accepted manuscript in the framework of the project.

To deposit your manuscript, please go to:

<http://peer.mpdل.mpg.de/helpdesk/wiki/PEERdeposit>.

For further information on PEER and guidance, please visit the PEER helpdesk:

<http://peer.mpdل.mpg.de/helpdesk>"

---

\* For the following analysis the PEER Project is indebted to MingXin Zhou, Adrian Mulligan and Mayur Amin at Elsevier, who processed and provided the data. The PEER Project would like to thank them for their help.

The PEER infrastructure offered a convenient and easy route to deposit through the following procedure:

- Authors received an invitation to deposit at the same time as the manuscript was accepted for publication, thus prompting deposit at a time when the manuscript was still fresh;
- As publishers sent out the invitation to deposit, this removed any possible concerns about permission and copyright, thus enabling authors to deposit with confidence;
- As the invitation specified the manuscript version to be deposited, and authors had this version at hand, this facilitated prompt deposit.

Principally, the robust infrastructure in conjunction with a convenient deposit mechanism enabled authors to self-archive. That they did so in small numbers only, raises the question if this can be explored any further?

Part of the context is that authors in some disciplines have been archiving their manuscripts online but, on the whole, they do not do so in large numbers. Concurrent with the spread of the idea of open access, surveys have indicated a growing awareness and principal willingness of many authors to self-archive their manuscripts (or have these self-archived).<sup>5</sup> Of course, support for an idea need not translate into actual behaviour. Despite a decade of organized open access and an explicit invitation from publishers, very few authors chose to self-archive within the PEER Project (though some may have self-archived in a different repository, which PEER was unable to track).

Also part of the context is that open access mandates seek to raise the deposit rate. The PEER Project emulated this scenario by having publishers deposit accepted manuscripts, thus ensuring that at least 50% of the eligible manuscripts became available. This situation corresponds well to many open access mandates, in which compliance rates vary between 30% and 70% (typically 40-50%) and this often hinges on publisher services (both open access publishing and manuscript deposit).<sup>6</sup>

As the PEER Project invited thousands of authors to self-archive, this would seem a good opportunity to explore the motivation of authors. However, firstly, the small number of authors that actually deposited impede this effort. Secondly, and principally, any effort to directly interview or survey the invited authors would intervene into the relationship between publisher (journal) and author, which is problematic legally and ethically. However, it is possible to analyze this specific cohort of authors by asking whether their observed behaviour correlates with any specific attributes such as discipline, country and number of publications.

---

<sup>5</sup> PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories Final Report. J. Fry et al., August 2011 [http://www.peerproject.eu/fileadmin/media/reports/PEER\\_D4\\_final\\_report\\_29SEPT11.pdf](http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf)

<sup>6</sup> Armbruster, C. (2010) Implementing Open Access Policy: First case studies. *Chinese Journal of Library and Information Science*, 3(4) 1-22

For the publication year 2010, based on Scopus data, it is possible to identify a cohort of 3,913 unique authors that had been invited to deposit an article in the PEER infrastructure. 118 authors deposited an accepted manuscript, while 3,795 did not. This is a deposit rate of 3%. Bearing in mind that this is a snapshot only (and some authors may have deposited elsewhere), it is noteworthy that junior and/or younger scholars were more likely to self-archive.

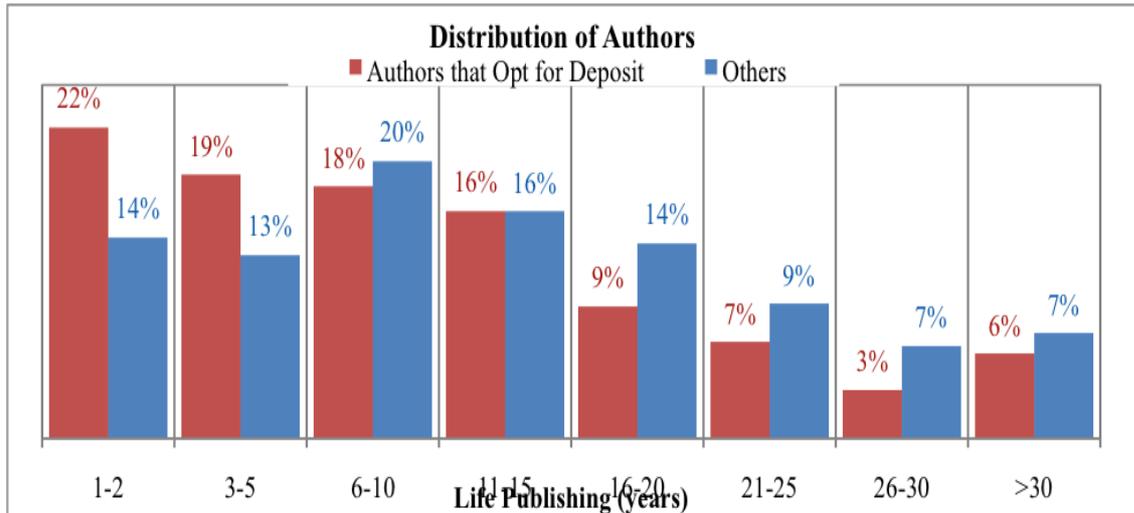


Fig 1. Distribution of authors depositing or not

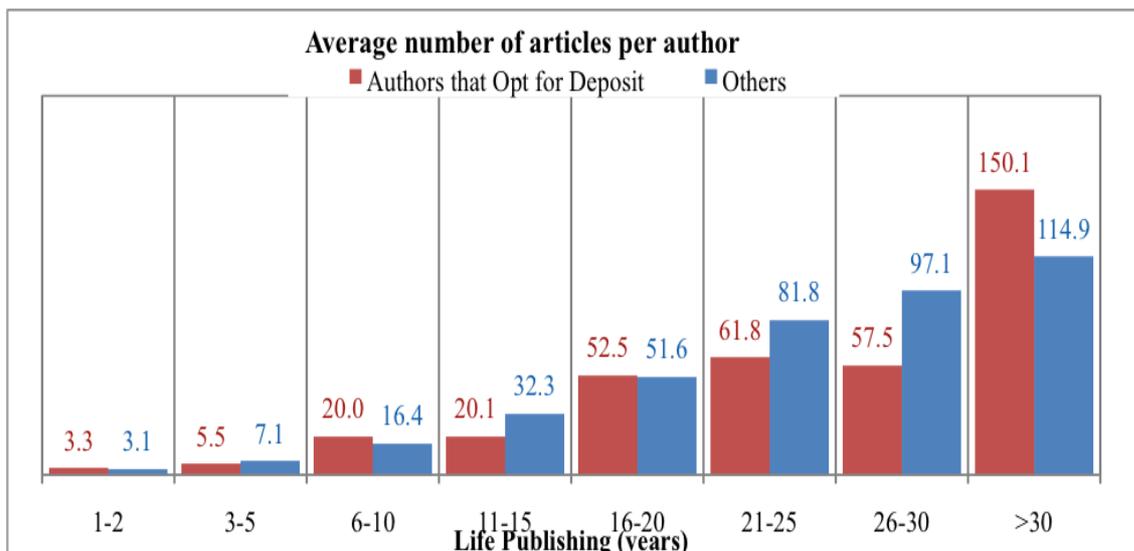


Fig 2. Average number of articles per author

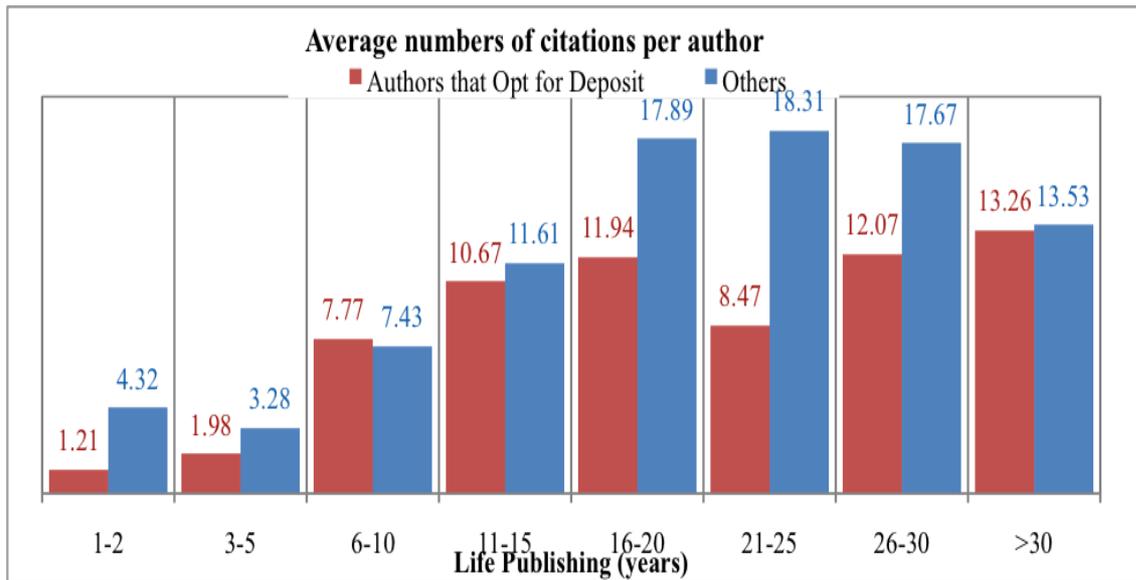


Fig 3. Average number of citations per author

Based on this data, the following observations may be made:

- The largest tier (22%) among authors depositing are those who most recently started publishing in scholarly journals, and the propensity to self-archive decreases with experience;
- The average number of articles published is lower for those authors choosing to deposit (28.5 articles) than for those not depositing (39.5), which seems a function of a shorter or longer publication history;
- The average number of citations for authors depositing is also lower than for those not (10.5 versus 14.8 citations), presumably a function of publication history and frequency.

The number of depositors is too small for any meaningful conclusions with regard to disciplines or countries. The finding that junior and/or younger scholars are more likely to self-archive should be tested across different scenarios.

In conclusion, the low number of authors willing to deposit in the PEER infrastructure is surprising given a growing awareness of open access and a convenient deposit mechanism. A preliminary analysis of the cohort indicates that junior and/or younger scholars were more likely to participate. Funders and institutions with an open access policy may be interested in further investigating the attitude and behaviour of authors invited or requested to self-archive.

## 6. User experience: Repository exit survey

The PEER Behavioural Research Team (Loughborough University, led by Jenny Fry) conducted a user survey via the PEER repositories [1]. Unfortunately, the response rate was very low. Hence, the PEER Project decided to try again. To attract more responses, the survey was implemented not just at the PEER repositories, but also across the repositories more generally. Generally, little is known about the expectations and behaviour of repository users and studying the use of Open Access Repositories is a complex matter. Hence, the user survey should be understood as a pilot project, exploring the potential for more research. The survey was implemented by the PEER partners and further project partners: six repositories in all. The analysis was undertaken by Heidemarie Hanekop (University of Göttingen).

**Summary:** We collected 299 responses from users visiting repositories – not a representative sample, but nevertheless a starting point for analysis. Most responses were collected at the HAL (archives ouvertes) and HAL INRIA national repositories, some from institutional repositories (MPDL, UGOE), and also from the subject-based repository SSOAR. At all repositories, more responses were obtained from users visiting the general collection than from those using the PEER collection. For more than 40% of the respondents Open Access repositories with free access to scientific articles is “essential”, and another 36% find them “very useful”. Usage patterns of institutional repositories on the one hand and the separate PEER repositories on the other are apparently quite different. Almost half of the users of the separate PEER repositories had not used an Open Access repository before, whereas three quarters of the users of the other participating repositories had visited an Open Access repository before. Furthermore, half of PEER users arrive at the repository searching for a specific article via Google. By contrast, almost half of the users of the other repositories went directly to the repository. Users of the separate PEER repositories do not use them for “current awareness”, but rather while “writing an article” or “for literature review”. Users at PEER repositories were sometimes confused when they wouldn’t find the Version of Record, but something else. 11% of the PEER users found the version of the article obtained “very satisfactory for their purpose”, compared to 34% of all respondents. However, the great majority did say that they would use an open access repository again when searching for an article.

### *Design and method of the survey*

The repository user survey (WP 7) was implemented at the repositories of HAL, INRIA, MPDL, UGOE and SSOAR. The questionnaire (see Appendix) was based broadly on the questionnaire used in the first study, but the present version is both shorter and more comprehensive. A link to the survey webpage was placed on each repository website, with the following invitation:

“Help us improve our Open Access Repository for published research articles. It will only take 5 minutes of your time, and you could win an Amazon voucher worth 50 Euros.”

A direct link to the online questionnaire was placed on the repository homepages at INRIA, HAL, MPDL, UGOE and SSOAR [3]. The PEER documents at INRIA, HAL, MPDL and

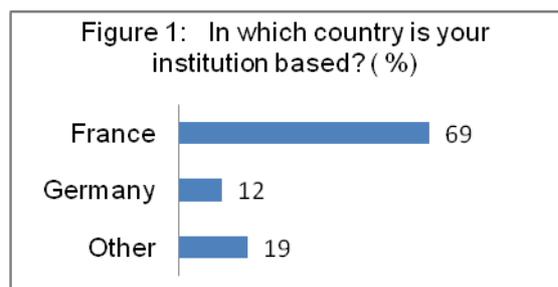
UGOE are hosted in separate repositories [4]. As the homepage of the separate repositories are not frequently used as a starting point, the link to the survey was placed on the search pages. Since we gathered the information about the repository visited by the respondent, we are able to compare PEER repository users to the users of other repositories.

The survey was implemented from October 2011 to the end of February 2012, with the exception of SSOAR, where the survey ran from the beginning of January to the end of March 2012. The link to the survey was clicked 1941 times, with 299 respondents completing the questionnaire. We may surmise some self-selection among respondents: those who have a positive attitude toward open access may be more likely to complete the questionnaire. Moreover, heavy users of repositories are likely to be over-represented in the sample. However, as this group has extensive experience, their responses are particularly helpful when conducting a pilot.

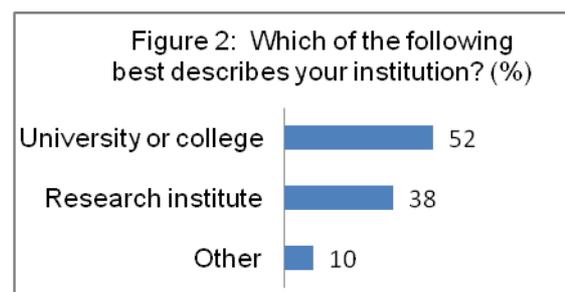
### *The Sample*

Most responses were collected at the HAL (archives ouvertes) and HAL INRIA national repositories, some from institutional repositories (MPDL, UGOE), and also from the subject repository SSOAR. At all repositories, more responses were obtained from users visiting the general collection than from those using the PEER collection. Three quarters of the respondents were visiting the national repositories of INRIA and HAL ( $n=171/55$ ), while 11% arrived from the institutional repositories of the University of Göttingen ( $n=5$ ) and the Max Planck Digital Library (MPDL) ( $n=27$ ). Another 5% of respondents were users of the Social Sciences Open Access repository (SSOAR) ( $n=16$ ) the subject-based repository in PEER. Eight percent ( $n=25$ ) of our respondents arrived from the separate PEER repositories of HAL/INRIA or MPDL or UGOE (see Fig. 5 and Appendix, Tab. 1).

Users of the French repositories are decidedly over-represented, as is also evident from the distribution of users by country: 69% come from France, 12% from Germany and 19% are from other countries (see Fig. 1).

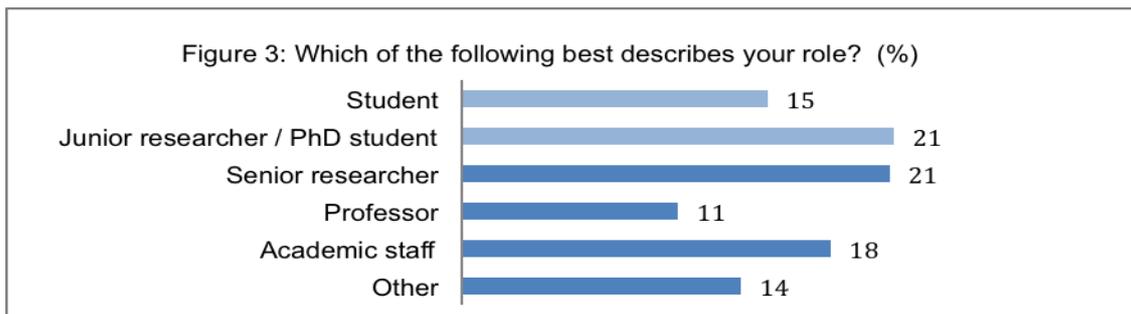


The latter are widely scattered, and include the US, the UK, and the Netherlands, as well as Eastern European countries, India, and others. Most of the respondents work at a university (52%) or a research institute (38%) (Fig. 2). As the distribution of respondents by nationality is skewed, this is considered when examining the differences between repository users.

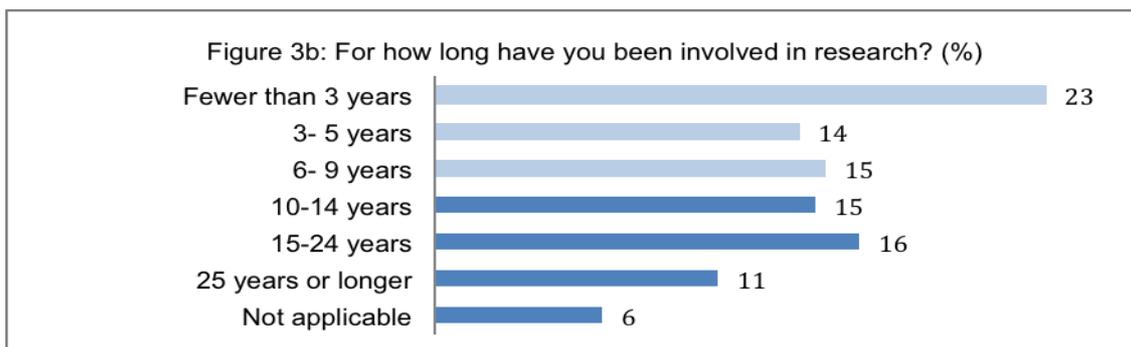


Disciplines of respondents are broadly distributed, with the highest percentages in “Physical sciences & mathematics” (30%) and “Social sciences, humanities & arts” (27%). The least well represented fields were Life sciences and Medicine (7%) (see Appendix, Tab. 2.1)

Compared to the first user survey [5], our sample has a more representative distribution of respondents by status and research experience. Senior researchers and professors are well represented, 22% of the respondents are senior researchers and as many as 11% are professors (Fig. 3); another 18% are academic staff members. One quarter are junior researchers or PhD students and not more than 15% are students.



More than 40% of respondents have been involved in research for over 10 years.



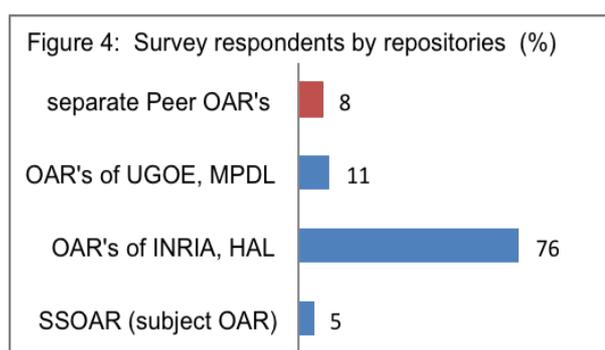
A majority of respondents are experienced and well-established researchers. This is also demonstrated by the fact that two-thirds of respondents had already published articles in academic journals.

Respondents are grouped by repository usage as follows (see Fig. 4):

- separate PEER repositories (at INRIA/HAL, MPDL, UGOE) with the PEER content
- institutional repositories of INRIA/HAL
- institutional repositories of MPDL and UGOE
- Social Science Open Access Repository (SSOAR), a subject-based repository

The separate PEER repositories all hold the same PEER content, supplied by the PEER Depot. They are designed so that articles are found via search engines, rather than via the repository home page.

The subject-based repository SSOAR is a special case insofar as the PEER manuscripts for this repository are integrated in SSOAR.



## Findings

### *Usage of all respondents*

Three quarters of all respondents had made use of an Open Access repository before (see Fig. 5 below, blue bar). Only 19% indicated that this was their first time using an open access repository. This corresponds to the fact that almost half of the users (44%) went directly to the repository website to search for articles (see Fig. 6 below, blue bar). In particular, the repositories of HAL and HAL INRIA are an important path of access to scientific publications for their users. One quarter of respondents arrive at the repositories through searches using Google and another 17% using Google Scholar (Fig. 6). Just as many arrived from a library portal (17%) or via a link from a website or reference (19%).

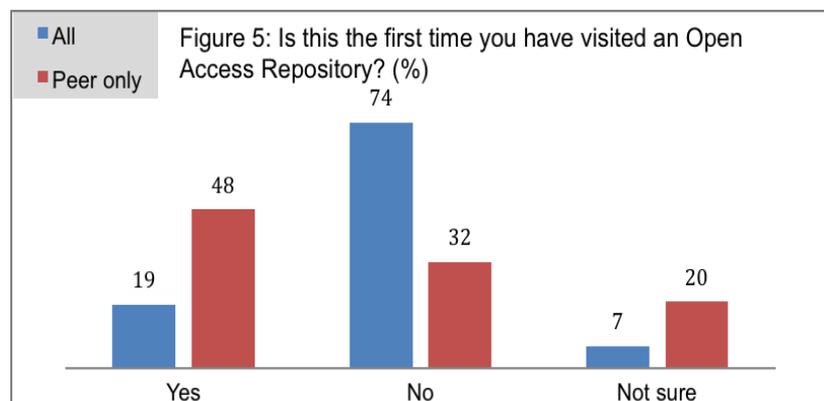
Almost half of the respondents in our sample (43%) indicated that open access repositories are “essential” for providing access to scientific papers. Another 36% find them “very useful”, 15% “useful”, and only 3% “not very useful” (Fig. 13 below, blue bar and Appendix Tab. 4.2). However, the majority of all respondents are regular users of open access repositories, and for them Open Access is very important.

For what purposes do the survey respondents visit the Open Access repositories? One-third use it “for current awareness and for keeping up to date” (Fig. 11, blue bar). This is somewhat surprising, since it is commonly assumed that institutional repositories are preferred for known item searches [6]. This may be due to the fact, that most respondents in this survey are regular user of the participating repositories and use them for several purposes. The most common purpose for usage an OAR is for “literature review” (Fig. 11, blue bar). Other purposes were also similarly important: “writing an article or report” (33%) or “exploring a new topic” (28%). Less important for our sample were “completing a student assignment” (8%) or “education and professional development” (15%).

Half of the respondents visiting the repository were looking for “articles about a particular topic (keyword search)”, but another 38% were searching for a specific article, and 35% for the work of a particular author or research group (Fig. 7, blue bar).

### *Usage of the separate PEER repositories compared with all respondents*

Usage of the separate PEER repositories [7] differs in many respects from the usage of the other repositories, while differences among users of institutional repositories (INRIA/HAL; UGOE and MPDL) are slight. While a majority of respondents had used open access repositories before, almost half of the PEER repository users (48%) were using an open access repository for the first time (Fig. 5).



In contrast to the users of the institutional repositories, only one-third of PEER users had previously used an open access repository (32%), another 20% were not sure. The majority of PEER users are not regular open access repository users; on the contrary, for half of them open access is a new experience.

Furthermore, only 20% of the users of the PEER repositories went directly to the PEER repository website, compared to 44% of all respondents. Half of the users of the PEER repositories arrive via Google (44%) or Google Scholar (8%); 32% arrived from another webpage (Fig. 6). PEER content is more likely to be accessed directly via a search engine than by visiting the homepage of the repository. Compared to the findings of the Final PEER behavioral research report, less users of the PEER repositories went directly to the repository, since the results for Google are very similar (50% of the respondents of the PEER behavioral study reached the repository via Google [8]).

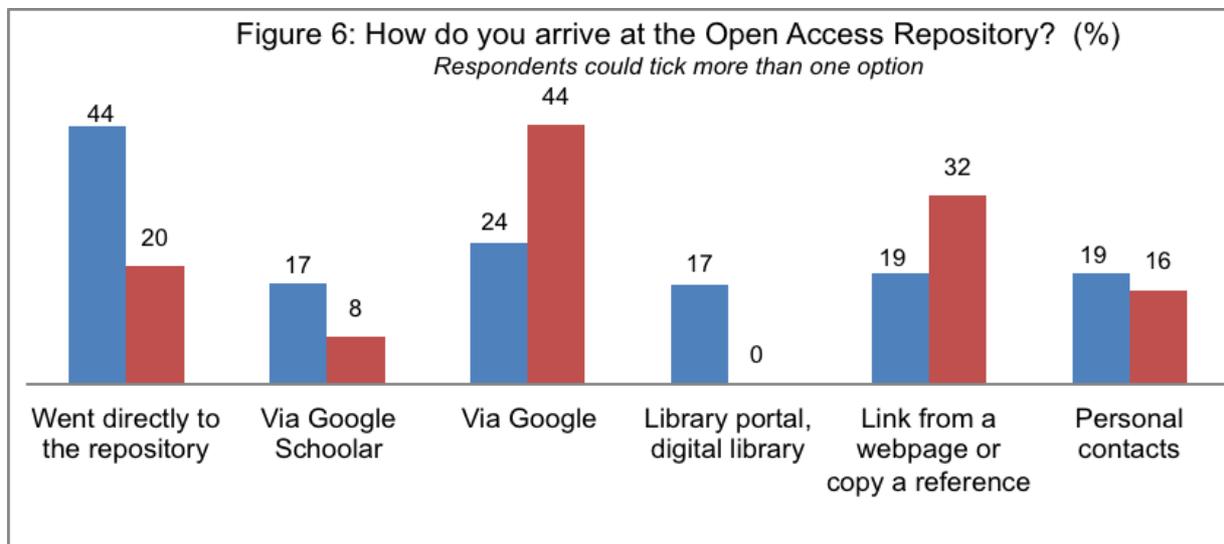
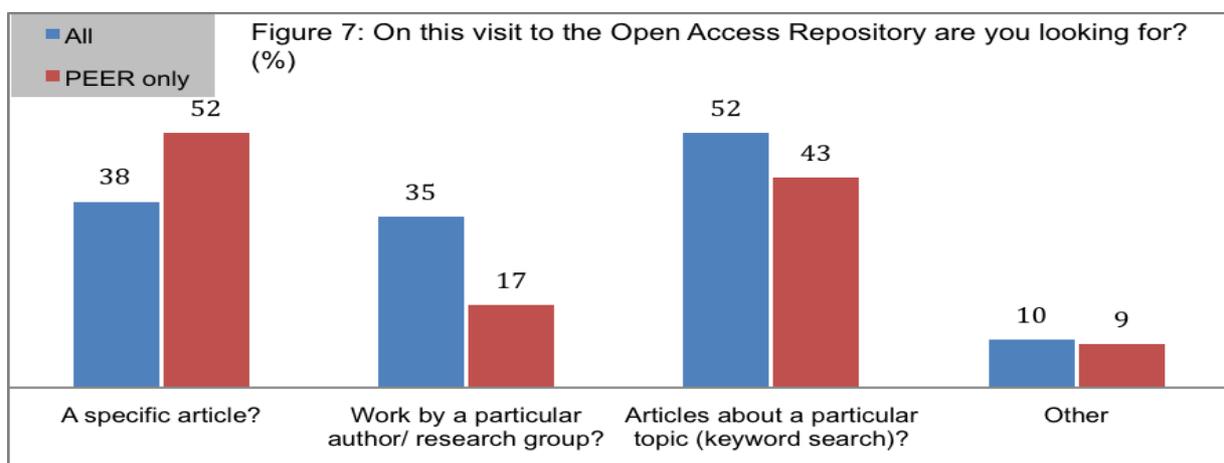


Figure 7 shows that 52% of PEER users (red bar) search for a specific article, while only 38% of all respondents (blue bar) said the same. On the other hand, only half as many PEER users as in the total group search for publications by a particular research group (17%).



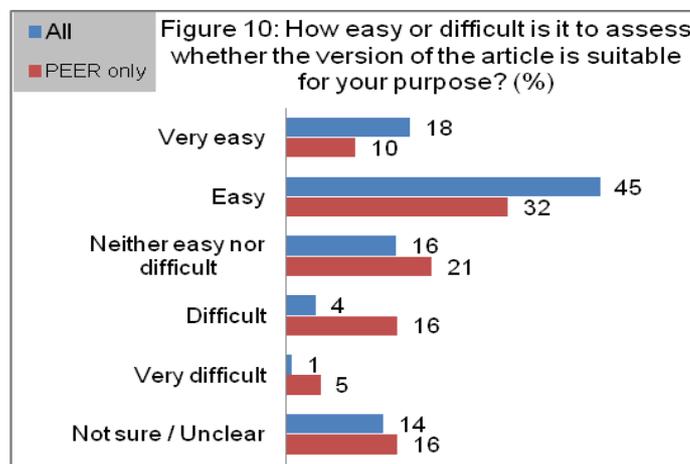
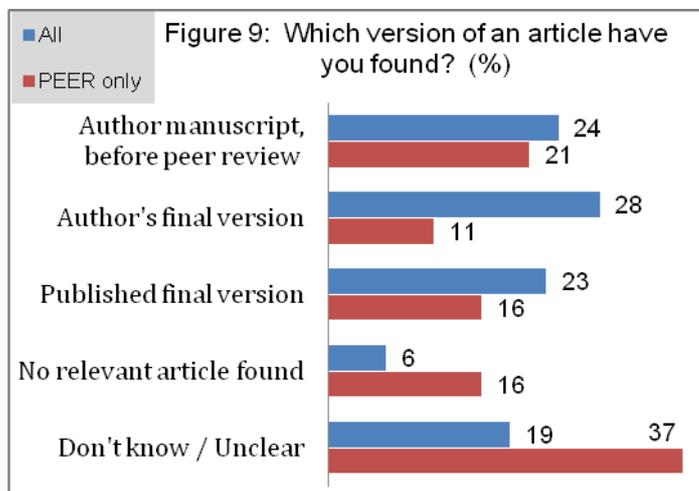
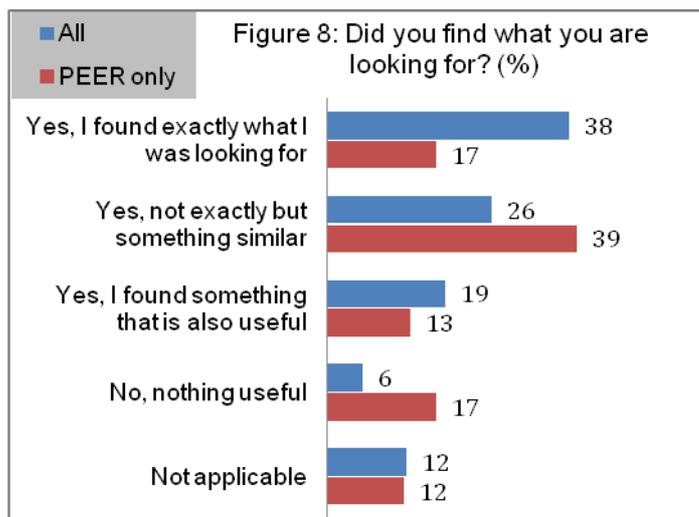
Did users of PEER repositories find what they were looking for?

Figure 8 shows that PEER users usually find what they are looking for, but in contrast to the total group, PEER users found less frequently exactly what they were looking for (17% compared to 38% of all respondents). But more PEER users found something similar (39% compared to 26% of all respondents). However, 17% indicate that they found “nothing useful”.

The problem seems to be that when users find a paper in a PEER repository, they often cannot decide which version it is. Figure 9 shows that 37% of PEER users (red bar) answered “don’t know/it’s unclear” to the question which version of an article they had found, compared to 19% of all respondents (blue bar). 16% of the PEER users (by contrast with 6% of the entire group) found “no relevant article”. Altogether, half of the PEER users reported problems with finding the relevant/right article, while this was true of only 25% of the users of the general repositories. A caveat is that the PEER collection, of course, is neither a subject-based nor an institutional collection.

42% of PEER users find it “easy” or “very easy” to assess whether the version of the article is suitable for their purpose, compared to 63% of all respondents. But more than one third (16% + 5% + 15%) of the PEER users find it difficult to determine whether the article found in a PEER repository is suitable for their purpose (Fig. 10).

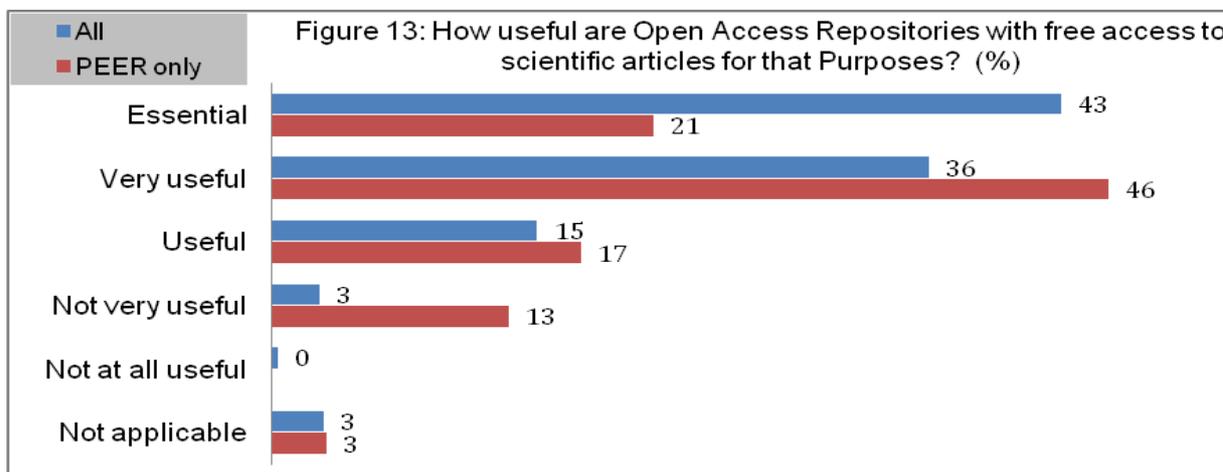
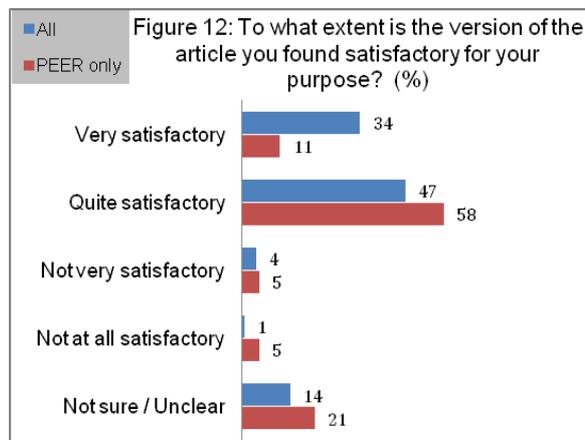
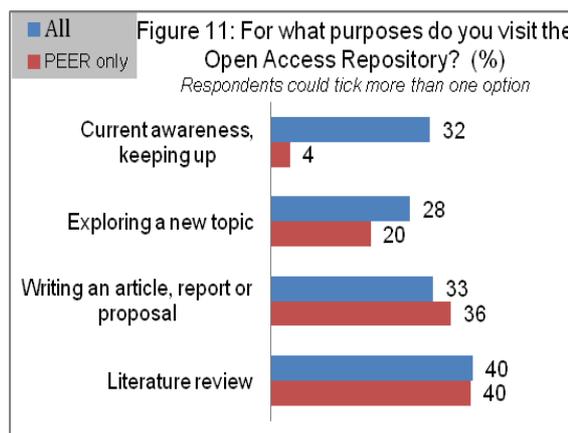
Also typical for the usage pattern of PEER users is that they don’t visit the PEER repository for “current awareness”, but for “writing an article” or “literature review” (Fig. 11). When the quality of the articles found by PEER users is unclear, this could reduce their usefulness for the two most important purposes for which the articles are sought: for writing and citing an article, report or proposal, and for literature review [9].



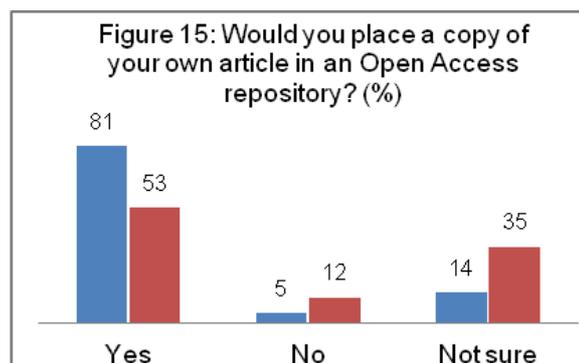
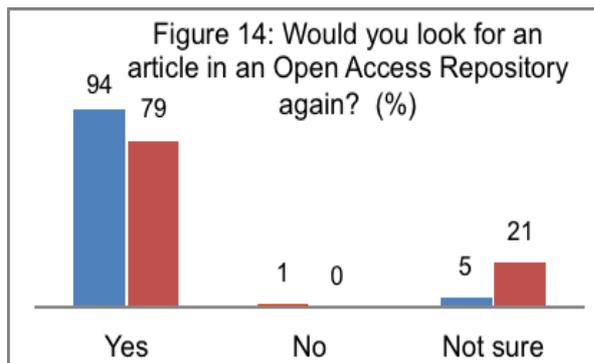
The vast majority of PEER users (11% + 58%) found that the article obtained was satisfactory for their purpose (Fig. 12).

However, PEER users are less satisfied with the version of the article they found than the majority of the respondents, since only 11% of PEER users think that the article they found is “very satisfactory” for their purpose – in contrast to 34% of all respondents. Compared to the Final PEER behavioural research report this is a lower rate of satisfaction. Whereas over half of the respondents in Behavioural Research indicated that the article they accessed was “quite satisfactory” for the task they were undertaking, with a further 40% indicating that the article was “very satisfactory”, only 6% said they “didn’t know/were not sure” [10].

Overall, PEER users find open access repositories with free access to scientific articles “very useful” (36% compared to 46%, Figure 13) or even “essential” (21%, compared to 43% of all respondents).

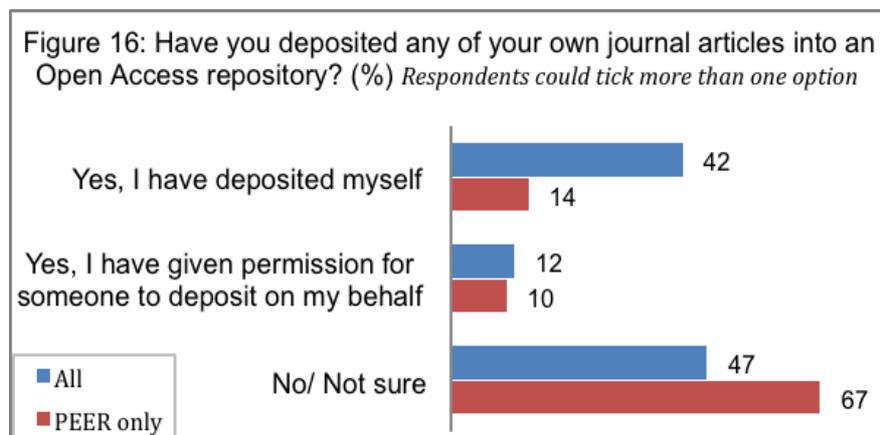


Though fewer PEER users “would look for an article in an open access repository again” (Fig. 14), nevertheless 79% would use an open access repository again.



More than half of PEER users (53%) said they would place a copy of their own article in an open access repository (Fig. 15), compared to 81% of all respondents. It is interesting to note, however, that the majority would place their own articles in an Open Access repository even though half of them for the first time used an Open Access repository.

The fact that PEER users are somewhat less familiar with open access is also seen in the fact that only 14% have deposited own journal articles in an open access repository (compared to 42% of all respondents; Fig. 16).



Our survey shows a difference in usage patterns between scientists, who routinely use Open Access repositories and users that search for an article they need for a specific purpose via Google (or other search-engines). Heavy open access users are familiar with green versions and Open Access repositories. They confidently assess whether the item found is fit-for-purpose. Users that are not familiar with Green Open Access or with repositories are more often less certain. Half of PEER users are new users.

### References:

- [1] PEER behavioural research Final report, p: V. and p: 83;  
[http://www.peerproject.eu/fileadmin/media/reports/PEER\\_D4\\_final\\_report\\_29SEPT11.pdf](http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf)
- [2] PEER behavioural research Final report p:12
- [3] <http://hal.archives-ouvertes.fr/>; <http://hal.inria.fr/>; <http://pubman.mpdl.mpg.de/>;  
<http://goedoc.uni-goettingen.de/goescholar>; <http://www.ssoar.info/>

- [4] INRIA/HAL: <http://peer.ccsd.cnrs.fr/>;  
MPDL: <http://peer.mpdل.mpg.de/pubman/faces/HomePage.jsp>;  
UGOE: <http://repository.peerproject.eu/jspui/>;
- [5] The first user survey had 34 respondents, where nearly 60% were students (35 % undergraduate or master, 24% PhDstudents) and 41% have been involved fewer than 3 years in research. (Final PEER behavioral research report p:102)  
[http://www.peerproject.eu/fileadmin/media/reports/PEER\\_D4\\_final\\_report\\_29SEPT11.pdf](http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf)
- [6] PEER Usage Research Interim report, CIBER, June 2010 p:7
- [7] PEER repositories are the separate PEER websites of INRIA/HAL, MPDL and UGOE in which only PEER articles are deposited  
(urls: [peer.ccsd.cnrs.fr/](http://peer.ccsd.cnrs.fr/); [repository.peerproject.eu/jspui/](http://repository.peerproject.eu/jspui/);  
[peer.mpdل.mpg.de/pubman/faces/HomePage.jsp](http://peer.mpdل.mpg.de/pubman/faces/HomePage.jsp))
- [8] Final PEER behavioral research report, p: 14  
[http://www.peerproject.eu/fileadmin/media/reports/PEER\\_D4\\_final\\_report\\_29SEPT11.pdf](http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf)
- [9] As the Final PEER behavioural research report also states; „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 final published version. These concerns are more prevalent where the purpose of reading is to produce a published journal article, and are perceived as less of an issue for other types of reading purpose.“ PEER behavioural research Final report, see p:73.  
[http://www.peerproject.eu/fileadmin/media/reports/PEER\\_D4\\_final\\_report\\_29SEPT11.pdf](http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf)
- [10] Final PEER behavioural research report, p. 16 and p. 100  
[http://www.peerproject.eu/fileadmin/media/reports/PEER\\_D4\\_final\\_report\\_29SEPT11.pdf](http://www.peerproject.eu/fileadmin/media/reports/PEER_D4_final_report_29SEPT11.pdf)

## Appendix

*The survey included the following questions:*

Page

- 1.1 Which of the following best describes your institution?
- 1.2 In which country is your institution based?
- 2.1 Which field best describes your research area?
- 2.2 Which of the following best describes your role?
- 2.3 For how long have you been involved in research?
- 3.1 Is this the first time you have visited an Open Access Repository?
- 3.2 How do you arrive at the Open Access Repository?
- 4.1 For what purposes do you visit the Open Access Repository?
- 4.2 How useful are Open Access Repositories with free access to scientific articles for that Purposes?
- 5.1 On this visit to the Open Access Repository are you looking for ..?
- 5.2 Did you find what you are looking for?
- 5.3 If you did not find what you are looking for, why not?
- 6.1 Which version of an article have you found?
- 6.2 To what extent is the version of the article you found satisfactory for your purpose?
- 6.3 How easy or difficult is it to assess whether the version of the article is suitable for your purpose?
- 7.1 Would you look for an article in an Open Access Repository again?
- 7.2 Do you publish articles in academic journals?
- 7.3 Would you place a copy of your own article in an Open Access Repository?
- 7.4 Have you deposited any of your own journal articles into an Open Access Repository?

***Tab. 1. Responses of users of participating Open Access repositories***

	Frequency	%
UGOE PEER	5	2
UGOE goescholar	5	2
MPDL PEER	18	6
MPDL pubman	27	9
INRIA	171	57
HAL	55	18
HAL PEER	2	1
SSOAR	16	5
Totals	299	100

### *Appendix: Repository User Survey responses*

Page 1

#### 1.1 Which of the following best describes your institution?

	Frequency	%
University or college	155	52
Research institute	113	38
Other		10
these are:		
Hospital or medical school	1	
Government	8	
Media	2	
Industrial/commercial	10	
Other	10	
Totals	299	100,0

#### 1.2 In which country is your institution based?

	Frequency	%
France	205	69
Germany	37	12
Other	56	19
Other	45	15,1
Netherlands	8	2,7
Austria	1	0,3
Belgium	1	0,3
Denmark	1	0,3
Totals	298	100,0

Page 2

## 2.1 Which field best describes your research area?

	Frequency	%
Medical sciences	9	3
Life sciences	12	4
Physical sciences & mathematics	87	30
Social sciences, humanities & arts	80	27
Interdisciplinary	35	12
Other	67	23
Not applicable	3	1
Totals	293	100

## 2.2 Which of the following best describes your role?

	Frequency	%
Student	44	15
Junior researcher / PhD student	62	21
Senior researcher	63	22
Professor	31	11
Academic staff	53	17
Other	40	14
Totals	293	100

## 2.3 For how long have you been involved in research?

	Frequency	%
Fewer than 3 years	69	23
3- 5 years	41	14
6- 9 years	44	15
10-14 years	43	15
15-24 years	48	16
25 years or longer	32	11
Not applicable	19	6
Totals	296	100

Page 3

## 3.1 Is this the first time you have visited an Open Access Repository?

	Frequency	%
Yes	56	19
No	218	74
Not sure	19	7
Totals	293	100

## 3.1 PEER-Repositories only:

Is this the first time you have visited an Open Access Repository?

	Frequency	%
Yes	12	48
No	8	32
Not sure	5	20
Totals	25	100

## 3.2 How do you arrive at the Open Access Repository?

	Frequency	%
Went directly to the repository	130	44
Via Google Scholar	51	17
Via Google	71	24
Library portal, digital library	50	17
Link from a webpage or copy a reference	56	19
Personal contacts	56	19
Other	13	4
Not sure	10	3
Via other search engine	14	5
Respondents could tick more than one option		

Page 4

## 4.1 For what purposes do you visit the Open Access Repository?

	Frequency	%
Current awareness, keeping up	89	32
Exploring a new topic	78	28
Writing an article, report or proposal	91	33
Literature review	113	40
Completing a student assignment	21	8
Professional development, education	43	15
Other	44	16

Respondents could tick more than one option

## 4.2 How useful are Open Access Repositories with free access to scientific articles for that purposes?

	Frequency	%
Essential	119	43
Very useful	99	36
Useful	40	15
Not very useful	7	3
Not at all useful	1	0
Not applicable	8	3
Totals	274	100

Page 5

## 5.1 On this visit to the Open Access Repository are you looking for?

	Frequenc	%
A specific article	100	38
Work by a particular author/ research	92	35
Articles about a particular topic (keyword)	137	52
Other	26	10

Respondents could tick more than one option

## 5.2 Did you find what you are looking for?

	Frequenc	%
Yes, I found exactly what I was looking	98	38
Yes, not exactly but something similar	66	26
Yes, I found something that is also useful	47	18
No, nothing useful	13	6
Not applicable	31	12
Totals	255	100

## 5.3 If you did not find what you are looking for, why not ...

	Frequenc	%
Author/ research group not found	21	8
Article not found or not suitable	39	15
Current article not found/ not up to date	24	9
Other	15	6

Not applicable	95	36
----------------	----	----

Respondents could tick more than one option

Page 6

### 6.1 Which version of an article have you found?

	Frequency	%
Author manuscript, before peer review	56	24
Author's final version	66	28
Published final version	53	23
No relevant article found	14	6
Don't know / Unclear	44	19
Totals	233	100,0

Respondents could tick more than one option

### 6.2 To what extent is the version of the article you found satisfactory for your purpose?

	Frequency	%
Very satisfactory	78	34
Quite satisfactory	108	47
Not very satisfactory	10	4
Not at all satisfactory	2	1
Not sure / Unclear	32	14
Totals	230	100,0

### 6.3 How easy or difficult is it to assess whether the version of the article is suitable for your purpose?

	Frequency	%
Very easy	42	18
Easy	106	45
Neither easy nor difficult	37	15
Difficult	10	4
Very difficult	2	1
Not sure / Unclear	33	14
Totals	234	100,0

Page 7

## 7.1 Would you look for an article in an Open Access Repository again?

	Frequency	%
Yes	221	94
No	3	1
Not sure	11	5
Totals	235	100,0

## 7.2 Do you publish articles in academic journals?

	Frequency	%
Yes	158	67
No	76	33
Totals	234	100,0

## 7.3 Would you place a copy of your own article in an Open Access Repository?

	Frequency	%
Yes	186	81
No	11	5
Not sure	32	14
Totals	229	100,0

## 7.4 Have you deposited any of your own journal articles into an Open Access Repository?

	Frequency	%
Yes, I have deposited myself	102	42
Yes, I have given permission for someone to deposit on my behalf	28	12
No/ Not sure	113	47
Respondents could tick more than one option		