Projekt PEER – Herausforderungen und Ergebnisse

Open Access Tage Wien, 27. September 2012
Barbara Kalumenos, STM

Supported by the EC eContentplus programme
Agenda

• Kurze Einführung

• Herausforderungen
  – PEER Beobachtungsraum

• Ergebnisse
  – drei Forschungsprojekte
  – Statements der Partner
Agreement and Disagreement

Agreement between publishing and research communities that access to results of publicly funded research is important to maximize its use and impact.

However, they hold different views on:
- whether mandated deposit in OA repositories is necessary
- the appropriate embargo periods
- impact on journal viability
Stakeholders in scholarly communication

- Publishers
- Researchers – authors and users
- Libraries and repositories
- Funding agencies

All of the above stakeholder groups are represented within PEER, both within the consortium & an advisory board
PEER: Background

- No clear evidence of effect of embargos
- STM proposes to HLG an experiment to find out

PEER starts September 2008
Current Situation

- Rapid growth of institutional repositories
- Individual funding agency mandates
- Publisher experimentation: allowing self-archiving
Current Situation

Number of Open Access Repositories

- Registry of Open Access Repositories
- Open DOAR

Publisher's allowing green Open Access policies (RoMEO)

- RoMEO green (Can archive pre-print and post-print)
- RoMEO blue (Can archive post-print (ie final draft post-refereeing))
- RoMEO yellow (Can archive pre-print (ie pre-refereeing))
- RoMEO white (Archiving not formally supported)
Project objectives

• PEER has been set up to monitor the effects of systematic archiving of ‘stage two’ research outputs (NISO: accepted manuscripts)

• Large-scale ‘experiment’ regarding deposit of author manuscripts: in an ‘observatory’ of OA repositories

• Research studies commissioned to gather hard evidence to inform future policies
  – Usage Research → Availability, usage
  – Behavioural Research → Author, reader behaviour
  – Economic Research → Costs, viability

• Collaborative project of diverse stakeholder groups
  – Publishers, research community and library/repository community
Project Overview

• Duration
  – 09/2008–05/2012 (3 years plus 9 months extension)

• Budget/Funding
  – 4.2 Mio €; 50 % by the European Union (eContentplus programme)

• Project partners
  – STM (coordination), ESF, UGOE, MPG/MPDL, INRIA
  – Technical partners: U. Bielefeld, SURF, KB Netherlands (long-term archiving)
  – 12 publishers
  – 6 repositories

• Contact / Website
  – peer@stm-assoc.org / http://www.peerproject.eu
Participating Publishers

- BMJ Publishing Group
- Cambridge University Press
- EDP Sciences
- Elsevier
- IOP Publishing
- Nature Publishing Group
- Oxford University Press
- Portland Press
- Sage Publications
- Springer
- Taylor & Francis Group
- Wiley-Blackwell
Participating repositories

- eSciDoc.PubMan.PEER, Max Planck Digital Library (MPDL), Max-Planck-Gesellschaft zur Förderung der Wissenschaften e. V. (MPG)
- HAL, CNRS & Institut National de Recherche en Informatique et en Automatique (Inria)
- Göttingen State and University Library (UGOE)
- SSOAR – Social Sciences Open Access repository (GESIS – Leibniz Institute for the Social Sciences)
- TARA – Trinity College Dublin (TCD)
- University Library of Debrecen (ULD)
- Long term preservation archive: e-depot, Koninklijke Bibliotheek
PEER Consortium

The PEER consortium (5 Executive members):

• International Association of Scientific, Technical and Medical Publishers (STM) - Co-ordinator
• European Science Foundation (ESF)
• Göttingen State and University Library (UGOE)
• Max Planck Gesellschaft (MPG)
• Institut National de Recherche en Informatique et en Automatique (INRIA)

Plus technical partners: SURF & Universität Bielefeld
Project Organisation
PEER Observatory

- The Observatory consists of
  - Publisher platforms (usage data & access to authors)
  - PEER Depot
  - PEER Repositories

- The PEER Depot
  - Acts as a „Clearing House“ - is a Dark Archive!
  - Processes deposits and distributes content to participating repositories

- The PEER Repositories
  - Provide the usage data (= log files) needed by our research partner CIBER

- Content inflow
  - 241 journals from four broad areas (Life Sciences, Medicine, Physical Sciences, Social Sciences & Humanities)
  - 2 ways of articles deposit: publisher deposit / author self-archiving
The PEER Observatory, content level & Research

- Publishers: 241 Eligible participating journals
- >53,000 mss
- Publishers submit 100% metadata
- Publishers invite authors
- Authors Self-deposit
- Central Deposit interface
- 11,800 invitations
- 170 mss
- Invited Europe based "PEER authors" to participate in survey for behavioural research
- Deliver usage data (log files) for usage research
- > 22,500 EU mss
- PEER DEPOT
- 100% EU manuscripts & metadata
- LTP:KB eDepot
- PEER REPOSITORIES
  - UGOE
  - SSOAR
  - MPG
  - HAL
  - ULD
  - TCD
- Embargo expired >18,000 mss

PEER – Publishing and the Ecology of European Research

www.peerproject.eu
Publisher deposits (cumulated)

Total amount of publisher provided content (~53,000 in October 2011)
PEER EU Deposits Processed (cumulated)

Author Deposits
Publishers Deposits
Embargo Expired

Author deposit 170
Embargo expired May 2012: ~19,000
PEER Observatory - Achievements

• Enormous efforts made and results obtained
  – A working large-scale Observatory which has delivered results!

• Functioning collaborative infrastructure

• Substantial quantities of content visible in repositories:
  ~19,000 EU deposits made publicly available (May 2012)
PEER Behavioural Research

Loughborough University, project leader: Jenny Fry

Two project phases (April – August 2009/ November – August 2011)

Conclusions (selected):

• ‘academic researchers do not desire fundamental changes in the way research is currently disseminated and published.’
• Researchers who associated Open Access with ‘self-archiving’ were in the minority (although this varies by discipline)
• authors tended to be favourable to Open Access but they do not want the pivotal role of the published journal article to be compromised
• Readers have concerns about the authority of article content and citability when the version they have accessed is not the published final version.
• Overall, repositories are perceived by researchers as complementary to, rather than replacing, current forums for disseminating and publishing research.
Two studies – Descriptive statistics (study 1) and randomised controlled trial (study 2)

Conclusions – Study 1:

„Limitations: …caution must be applied to the findings of this study. We absolutely should not generalize from the findings here to green open access more generally since PEER has a number of characteristics…… „ (p.20)

- FT Downloads are growing in a linear cumulative fashion for PEER and publishers however publishers are growing at a faster rate
- Relative popularity of PEER reveals considerable variation between publishers for reasons that are not yet clear.
- PEER content is SSH and physical sciences is significantly more popular than content in medicine, life sciences
Conclusions – study 1 (cont.):

• Analysis of cumulated FT downloads by age of article shows that articles continue to accumulate over a long period (an 18th month window represent only small proportion of lifetime).

• Substantial content arrived PEER during 2011. This makes is therefore difficult to interpret the findings in relation to embargo periods

• Analysis of publisher:repository downloads shows that users tend to prefer the publisher site for more recent content

• Article-level usage correlates positively and significantly across the publisher-repository divide. Articles popular on the one, tend also to be popular on the other, but correlation coefficients are modest.
Conclusions – Study 2:

- Exposure of articles in PEER repositories is associated with an uplift in downloads at the publishers´ web sites. Likely result of quality PEER metadata, a liberal attitude towards allowing search engine robots to index and higher digital visibility that PEER creates for scholarly content.

- Statistically significant was the positive effect only in the life and physical sciences.

- Larger publishers experienced stronger uplift; increase for smaller publishers was much weaker.

"The overall conclusion of this study is that there is no experimental evidence to support the assertion that PEER repositories negatively impact publisher downloads. Further research is recommended.." (p5).
A series of case studies, the Economics team explored costs drivers for publishers and repositories.

Findings:

• Cost ranges for peer review (which has no economies of scale)
• production activities and platform maintenance costs were obtained for publishers.
• repositories may have large sunk costs that are not accounted.
• They anticipate that publishers (subscription and Open Access) and repositories will increasingly be affected by ‘sustainability and competition for resources and reputation’.
PEER Executive Partners – Achievements & Reflections
Points of Agreement - PEER Executive Partners

- Building a large-scale infrastructure is organizationally and technically challenging
- Building a clearing-house with automated workflows is helpful
- Author self-archiving is unlikely to generate a critical mass of Green OA content.
- Stage II (accepted manuscript) archiving requires manual oversight and intervention
- Scholars prefer the Version of Record (indicated by the behavioural research as well as usage log analysis)
- Usage scenarios for Green Open Access are more complex than generally acknowledged
- The acceptance and utility of open access publishing has increased rapidly
Vielen Dank für Ihre Aufmerksamkeit!

FRAGEN ?

Alle Reports, Statements, Aufzeichnungen.... der PEER Konferenz in Brüssel vom Mai 2012 sind auf der PEER Webseite zu finden:

http://www.peerproject.eu
What is a Stage 2 manuscript?

- **Stage One** (NISO Author’s original)
  - Primary Outputs of Research:
    - raw data
    - Draft for submission to a journal

- **Stage Two** (NISO Accepted Manuscript)
  - Author’s manuscript incorporating peer review enhancements & as accepted for publication

- **Stage Three** (NISO Version of Record)
  - Final published article on journal website: version of record with copyediting, typesetting, full citability, cross-referencing, interlinking with other articles, supplementary data
PEER Depot Workflow (what goes on in the black box)

Publishers

- Articles
  - Metadata for publisher submitted articles
  - Metadata for author submitted articles

Authors

- Articles

PEER Depot

- All publisher submitted articles
  - Filtering: Journal? Article type? EU author?
  - Rejected deposits
  - "Selected articles"
    - Metadata matching: doi + pubdate available?
      - Metadata incomplete
        - pass2 received
        - Under embargo
        - embargo expiry
      - Metadata complete
        - Embargo expired
        - Article transfer to repositories & LTP depot
  - Metadata complete
  - Embargo expired
  - Article transfer to repositories & LTP depot

- All author submitted articles
  - Matching with publisher provided metadata. Journal? Article type? EU author?
  - Rejected deposits
  - "Selected articles"
    - doi + pubdate available?
      - Metadata incomplete
        - pass2 received
        - Metadata in complete
      - Metadata complete
        - Embargo expired
        - embargo expiry
        - Under embargo
        - Article transfer to repositories & LTP depot

GroBID – metadata extraction

Metadata → TEI

Metadata matching: doi + pubdate available?

Article transfer to repositories & LTP depot