

**peer**

Publishing and  
the Ecology of  
European  
Research

**Dr Julia M Wallace**

**Project Manager – PEER**

**julia@iglooe.com**

**STM Frankfurt Conference**

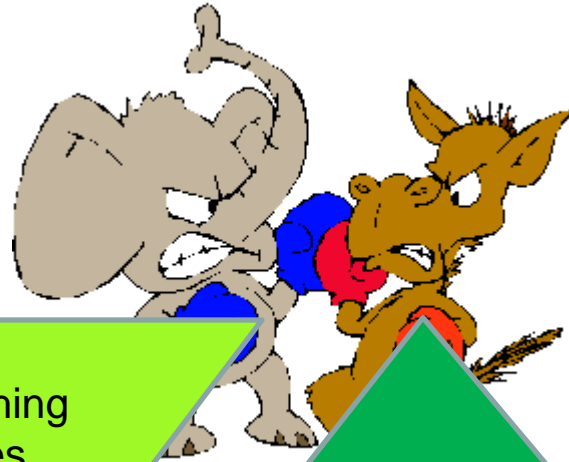
**9 October 2012**

# **PEER Project Update**



Supported by the EC eContent*plus* programme

# 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

# Project objectives

- **PEER was 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

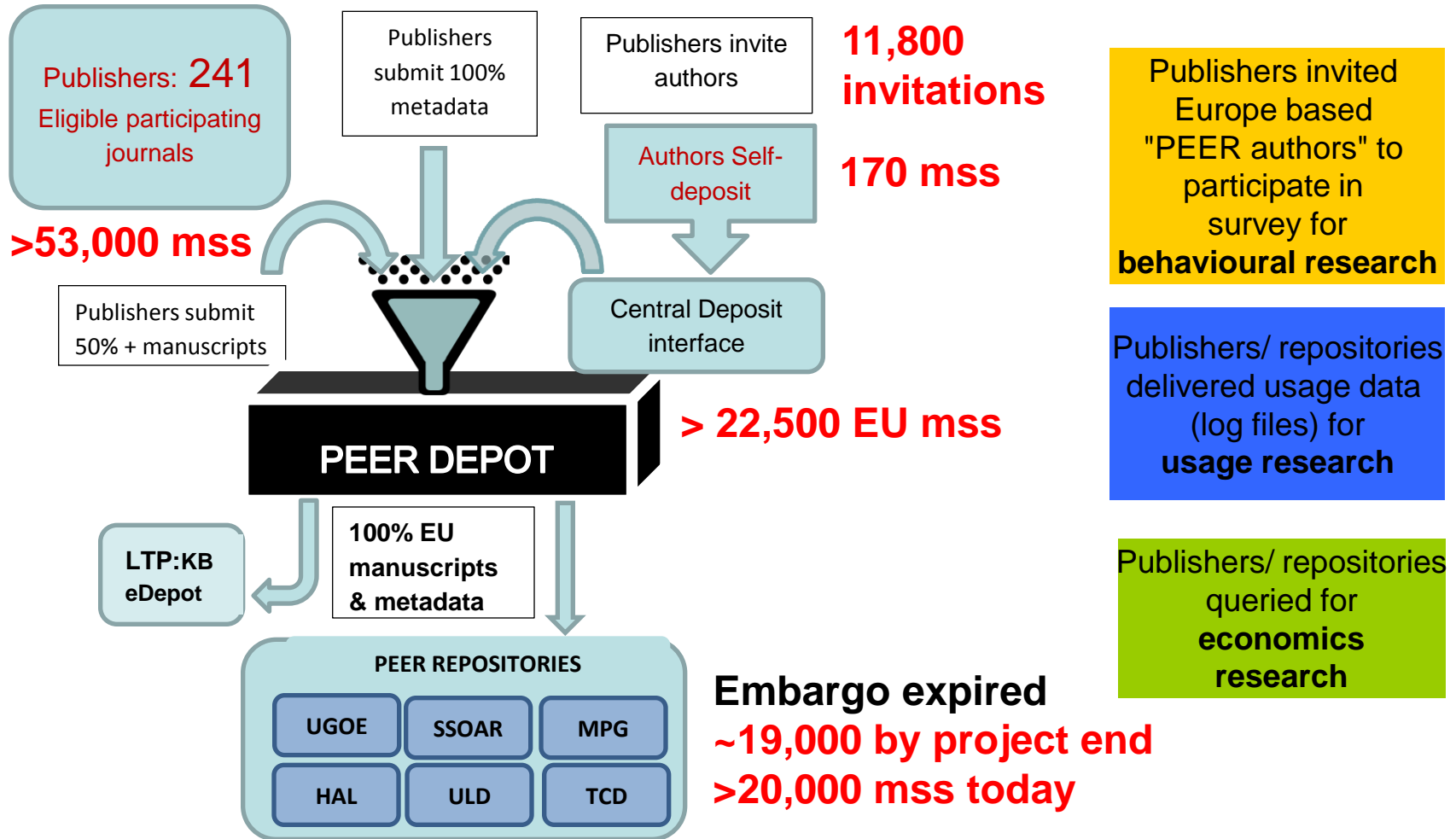
# Project Overview

- Duration
  - 09/2008–05/2012 (3 years plus 9 months extension)
- Budget/Funding
  - €4.2M : 50% from the EC (eContent*plus* programme) 50% partners
- PEER by Numbers
  - 5 Partners: STM (coordination), ESF, UGOE/SUB, MPG/MPDL, Inria
  - 2 Technical partners: Uni Bielefeld, SURF
  - 12 Publishers ; 241 participating Journals
  - 1 Depot/ Dark Archive
  - 6 Repositories
  - 1 Long-term preservation archive
  - 3 Research studies
  - 18 Advisory Board Members

# Participating Publishers, Repositories & LTP archive



# The PEER Observatory & content levels



# PEER Challenges and Solutions (1)

## PUBLISHER CHALLENGES

- **Stage two (accepted manuscripts) not standard extraction point**
- **Author accepted manuscripts in a variety of file formats**
- **All article types submitted**
- **Metadata delivery in several batches**
  - Article metadata are incomplete at acceptance time; Publication date unknown, DOI not attributed
  - Extraction of only „EU“ authored manuscripts not possible at acceptance stage
- **Different metadata formats**
  - NLM2.x, NLM 3.0, ScholarOne, proprietary
- **Some Metadata elements delivered within PDF document**

## PUBLISHER / PEER DEPOT SOLUTIONS

- **Change Process at Publishers**
- **Only one file format allowed – PDF**
- **Checking mechanisms: journal/article**
  - ISSN check
  - article type check
- **Article kept until metadata completion**
  - Metadata are accepted in either one step (on publication) or two passes (on acceptance and on publication)
  - EU author filter done at PEER Depot
- **Mapped onto single TEI structure**
- **Extraction done at PEER Depot (GroBID) in order to increase content**

# PEER Challenges and Solutions (2)

## REPOSITORY CHALLENGES

- Varying metadata requirements
- Varying ingestion processes
- Hosting PEER content
- Not configured for accurate embargo management
- Author authentication
- Logfile provision

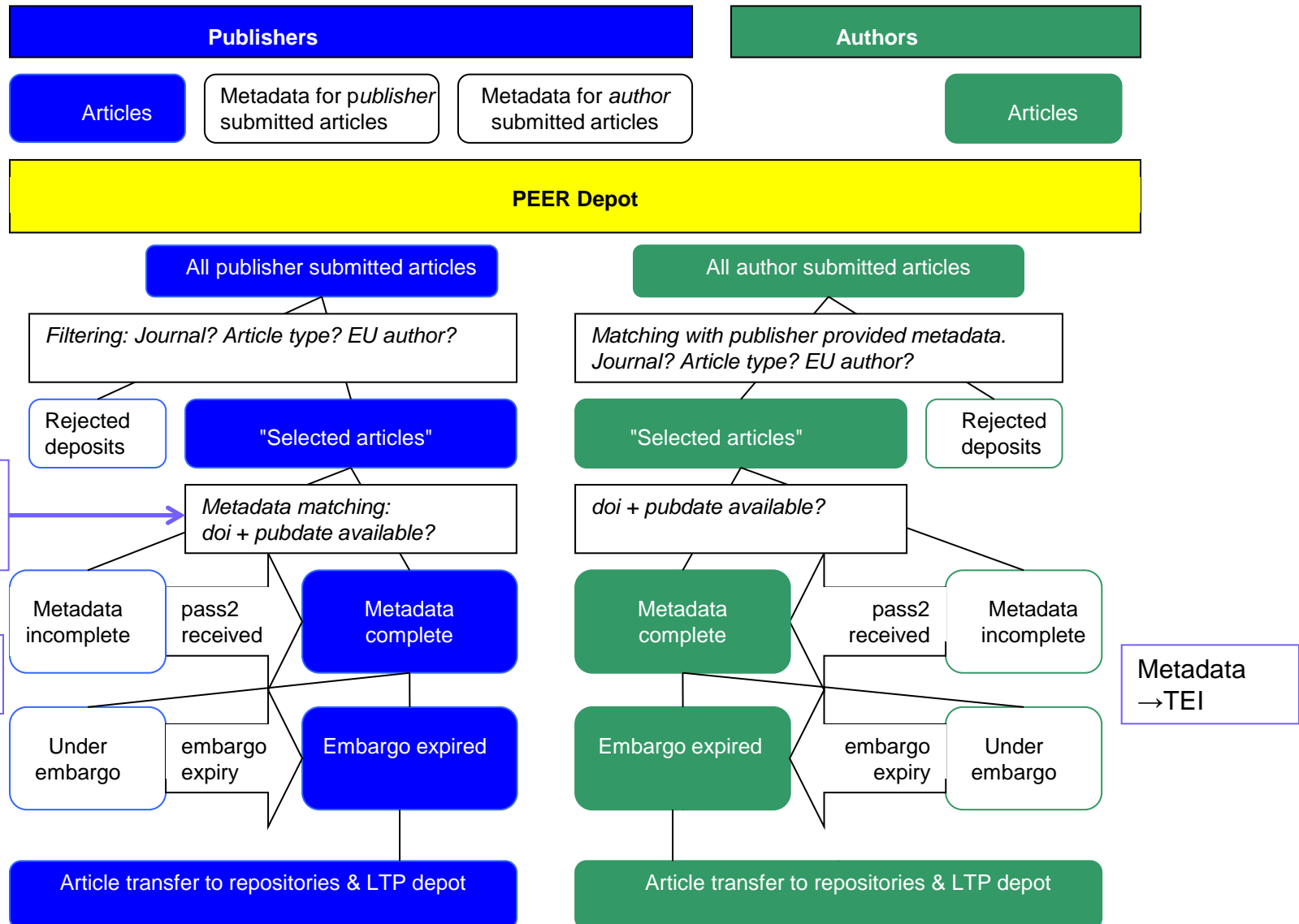
## REPOSITORY / PEER DEPOT SOLUTIONS

- Convert TEI metadata into internally used metadata standard
- Implement SWORD protocol for transfer between Depot & repositories
- Build dedicated PEER Repository within framework of home institution
- Embargo management undertaken at PEER Depot (0 - 36 months)
- Central deposit interface at MPDL then transfer to PEER Depot
- Set up anonymisation process plus automated transfer to Usage team

Other issues: Format and content problems with legacy manuscripts; Technical & financial challenges for repository participation (non PEER Partner repositories)



# PEER Depot Workflow (what goes on in the black box)



# PEER Research Projects

**Usage research • Behavioural research • Economics research**

High-quality, credible research, neutral, transparent and supported by all stakeholder groups

## **Research Oversight Group (ROG)**

Expert panel comprising three independent experts in scholarly publications and economics research:

- Carol Tenopir, University of Tennessee (USA)
- Cherifa Boukacem, Lille University (France)
- Tomàs Baiget, *El profesional de la Información*, Barcelona (Spain)

**Plus Industry advisor:** Mayur Amin, Elsevier

- Validate the specification for the research
- Advise on methodologies
- Evaluate the deliverables and confirm that the data is sound and conclusions are valid

# Behavioural research: Final Report

**Department of Information Science and LISU at Loughborough University, UK**

- Two phases of Research between 2009 and 2011
  - adopted a mixed methods approach consisting of surveys, focus groups and an interdisciplinary workshop

## SOME KEY CONCLUSIONS

- **Researchers who associated Open Access with ‘self archiving’ were in the minority.**
- Open Access is more likely to be associated with ‘self archiving’ (Green Road) by researchers in the Physical sciences & mathematics and the Social sciences, humanities & arts, than those in the Life sciences and the Medical sciences who are more likely to associate Open Access with Open Access Journals (Gold Road).
- **There is anecdotal evidence that some researchers consider making journal articles accessible via Open Access to be beyond their remit.**

# Behavioural research: Final Report

SOME KEY CONCLUSIONS continued:

- **Authors tend to be favourable to Open Access** and receptive to the benefits of self-archiving in terms of greater readership and wider dissemination of their research, **with the caveat that self archiving does not compromise the pivotal role of the published journal article.**
- **Readers have concerns about the authority of article content and the extent to which it can be cited when the version they have accessed is not the published final version.** 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.
- **Academic researchers** have a conservative set of attitudes, perceptions and behaviours towards the scholarly communication system and **do not desire fundamental changes in the way research is currently disseminated and published.**
- **Open Access Repositories** are perceived by researchers as complementary to, rather than replacing, current forums for disseminating and publishing research.

# PEER Usage Research

**CIBER Research Ltd, UK**

- High volume of content in the project (>**18,000 EU deposits** publicly available March 2012) supporting research with a high degree of confidence
- Measure activity over 12 months, starting March 2011
- Log file collection & analysis 1 March - 31 August 2011
- Randomised Controlled Trial: suppression of 50% PEER content at partner repositories. Logfile collection & analysis 1 December 2011 – 29 Feb 2012

Caveats:

The 'PEER' system is unlikely to have reached 'steady state' when study undertaken

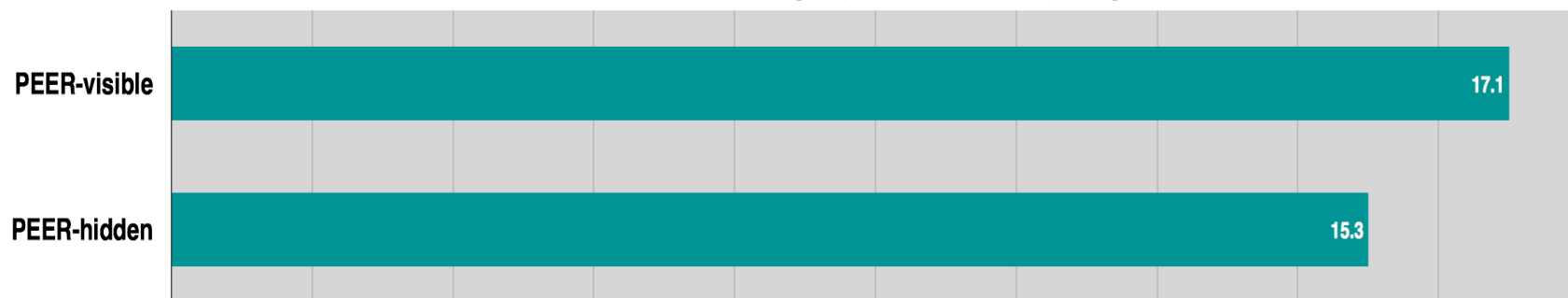
Observations of effects within PEER cannot be extrapolated to the Open Access universe

# ‘No effect’ publisher hypothesis: key findings

**Making preprints visible in PEER is associated with more traffic to the publisher sites.** (significant, if relatively modest increase)

Publisher full text downloads increased by 11.4%

95% confidence intervals: 7.5% to 15.5%, highly statistically significant at  $p < 0.01$



Publisher downloads went up in all subject areas, but with variation:

**Statistically *significant* increases in life sciences: up 20.3%** (13.1% to 27.9%,  $p < 0.01$ ) ; **physical sciences: up 13.1%** (5.2% to 21.6%,  $p < 0.01$ )

**Statistically *insignificant* findings in medicine: up 5.2%** (-1.0% to 11.7%,  $p = 0.10$ ) ; **social sciences and humanities: up 4.1%** (-0.05% to 13.9%,  $p = 0.38$ )

# PEER Usage Research: Conclusions

- The likely mechanism is that PEER offers high quality metadata, allows a wider range of search engine robots to index its content than the typical publisher, and thus helps to raise the digital visibility of scholarly content. There are variations as we zoom in on the detail and the jury is still out in medicine, the social sciences and humanities, and for smaller publishers, for reasons we do not understand yet.
- Publisher downloads are growing at a faster rate than PEER downloads and unless there is a step change, PEER's share of the market is likely to decline gradually over time.
- **What this research tells us is that the scholarly web is a complex environment, one in which digital visibility is king. Researchers make little use of the search facilities on repository or publisher sites, relying heavily instead on third-party gateways and general search engines.**

# PEER Economics Research: Conclusions

**ASK research centre, Bocconi University, Milan, Italy**

First detailed empirical study of cost drivers of publishers and repositories (22 organizations)

- SB and OA journals increasingly in competition for reputation and service;
- OA Article charges will become a reference for price setting
- Increased competition between big and marginal players;
- Increased competition among OA journals on reputation and article processing charges
- The characteristics of the digital platform key
- Successful digital journals increase their scope and broaden the disciplines covered



# The cost structure of journals

Cost category	Activities	Average direct cost per article	Cost drivers
Content certification	Management of peer review	250 USD	Number of articles per journal Journal rejection rate Number of reviewers per manuscript Number of rounds of review
Content publication (production)	Formatting Editing Typesetting Metadata	170 - 400 USD	Number of articles per journal Number of issues per journal Externalisation Standard cost of labor

- Cost of content management:
  - Investment in digital platform                      from USD 1 million up
  - Maintenance costs                                              USD 170 k– 400k

# Repositories: A lean cost structure ....

- Cost of processing documents (including metadata creation)
  - 10 EUR max per reference
  - 18 EUR max per full text
  - 43 EUR max per journal article
- Positive impact of harvesting and mandates in the speed up of feeding process
- Set up cost of repository was not determined;
- Cost of technical FTE per item
  - between 2 and 50 EUR per reference
  - between 2,5 and 53,2 per full text journal article

**.... with a big impact of sunk and organizational costs**

# Publisher points-of-view (1)

- All publishers emphasize their OA publishing ventures and their hybrid offers
- Collaboration with the PEER Depot was appreciated – for some the investment was significant, others already had complementary workflows and/or software
- Some publishers have few or no issue(s) with manuscript self-archiving, but the PEER Project confirms that publisher cooperation would be required for any large-scale Green OA scenario – and this would be a game changer

## Publisher points-of-view (2)

- If publishers are part of a Green OA scenario (& not all can imagine this), then their expertise, cost and opportunity costs need to be recognized
- If libraries strive to introduce usage as a (additional) pricing mechanism, then the proliferation of additional manuscript versions (e.g. Stage I, II) via repositories is highly problematic
- Given the growth of open access publishing (gold & hybrid) & the investment of publishers in archiving, are the (substantial) additional costs for Green OA justified?

# Repository opinion (1) – Deposit model

## **Systematic provision of a large number of manuscripts by publishers**

- enables automated deposit of manuscripts and metadata
- expands the number of journal articles available to various OARs
- *some repositories consider as part of future green OA policies*

## **Deposit of stage two manuscript versions**

- Heterogeneous versions from publishers, some originally not intended for public distribution
- quality check or standardization procedure would be desirable (by OAR's central depot or publishers)?

# Repository opinion (2) – Transfer process

## Efficient, automated transfer process

- development, implementation and intensive practical testing
- large quantities of documents and metadata to be processed and disseminated to repositories
- SWORD protocol for exchange of documents and metadata
- *some repositories consider as part future green OA policies*

## Central processing unit (PEER depot)

- clearing house, filtering EU-authors
- matching manuscripts with metadata
- embargo management
- *some repositories consider as part of future green OA policies*

# Executive Partners- Points of Agreement

- **Building a large-scale infrastructure is organizationally and technically challenging — even at a project level**
- **Author self-archiving is unlikely to generate a critical mass of Green OA content**
- **Version II archiving requires considerable manual oversight and intervention**
- **Scholars prefer the Version of Record**
- **Usage scenarios for Green Open Access are more complex than generally acknowledged**

# Executive Partners- Points of Agreement

- **The acceptance and utility of open access publishing (“Gold”) has increased rapidly**
- **A successful collaboration for experimental results**
- **Mutual understanding and trust**
  - Working together to manage and deliver the project encouraged professional respect on all sides, particularly in challenging or difficult moments



**For further information on PEER including final research reports and presentations made at the PEER End of Project Conference visit:**

**[www.peerproject.eu](http://www.peerproject.eu)**