



**Dr Julia M Wallace**

**Project Manager – PEER**

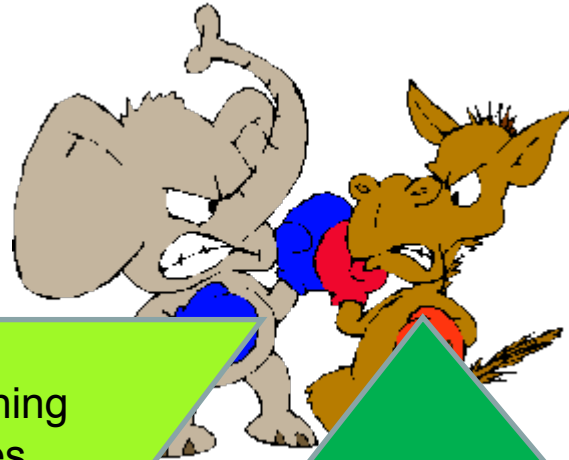
**wallace@stm-assoc.org**

**UKSG, 26-28 March 2012**

**Glasgow, UK**

# **The PEER Project: Investigating the Effects of Green Open Access**

# 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

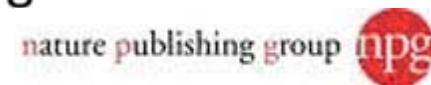
- **PEER has been set up to monitor the effects of systematic archiving of ‘stage two’ research outputs (NISO: accepted manuscripts)**
- Large-scale ‘experiment’ regarding deposit of author manuscripts: in an ‘observatory’ of OA repositories
- **Research studies commissioned to gather hard evidence to 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
  - €4.2M : 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

# 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)



Social Science Open Access Repository

- TARA – Trinity College Dublin (TCD)



- University Library of Debrecen (ULD)



- *Long term preservation archive*: e-depot, Koninklijke Bibliotheek

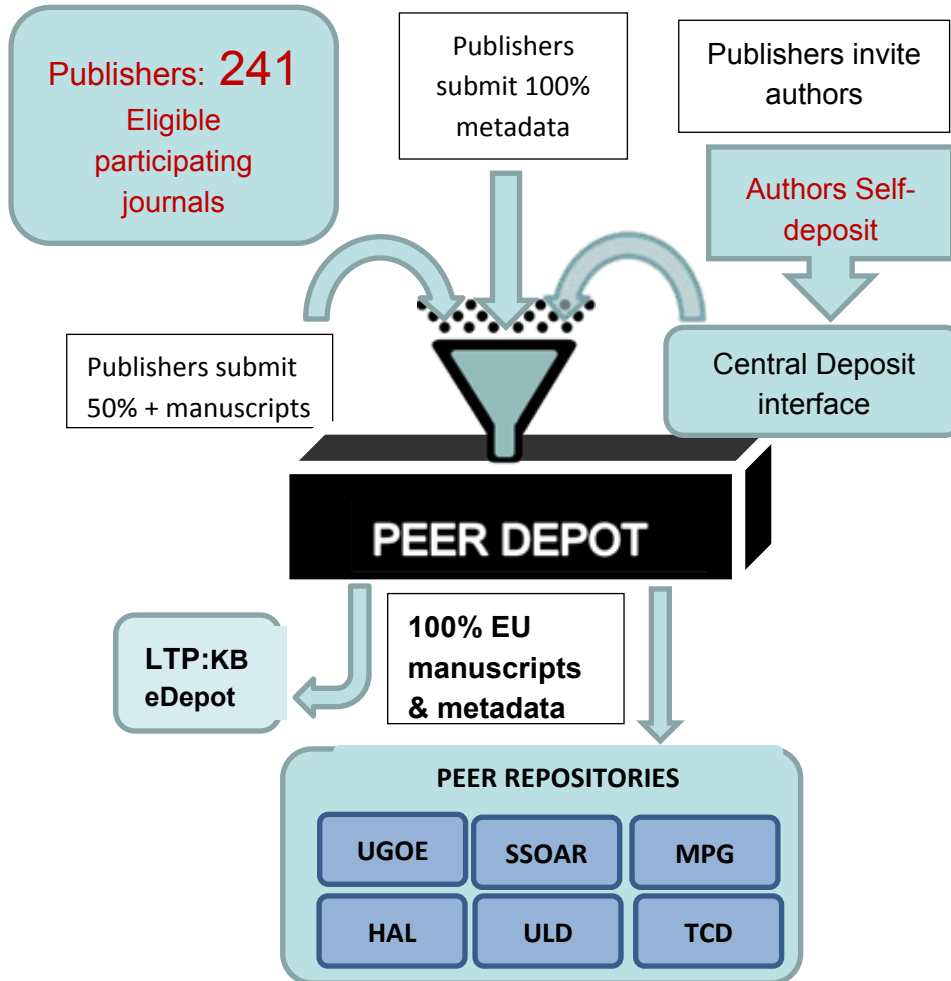


**KB** Koninklijke Bibliotheek  
Nationale bibliotheek van Nederland

# 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

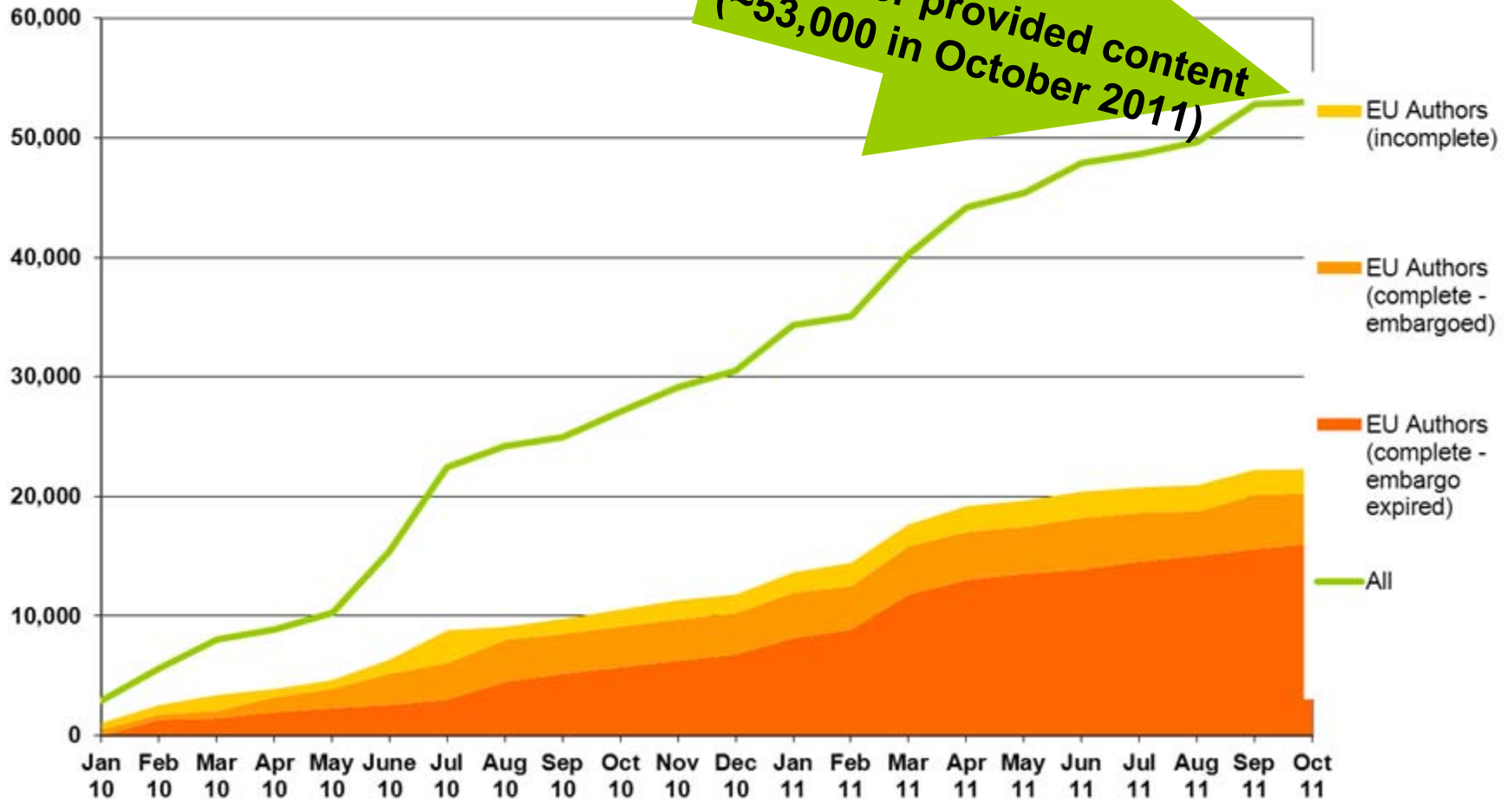
# The PEER Observatory – content flow



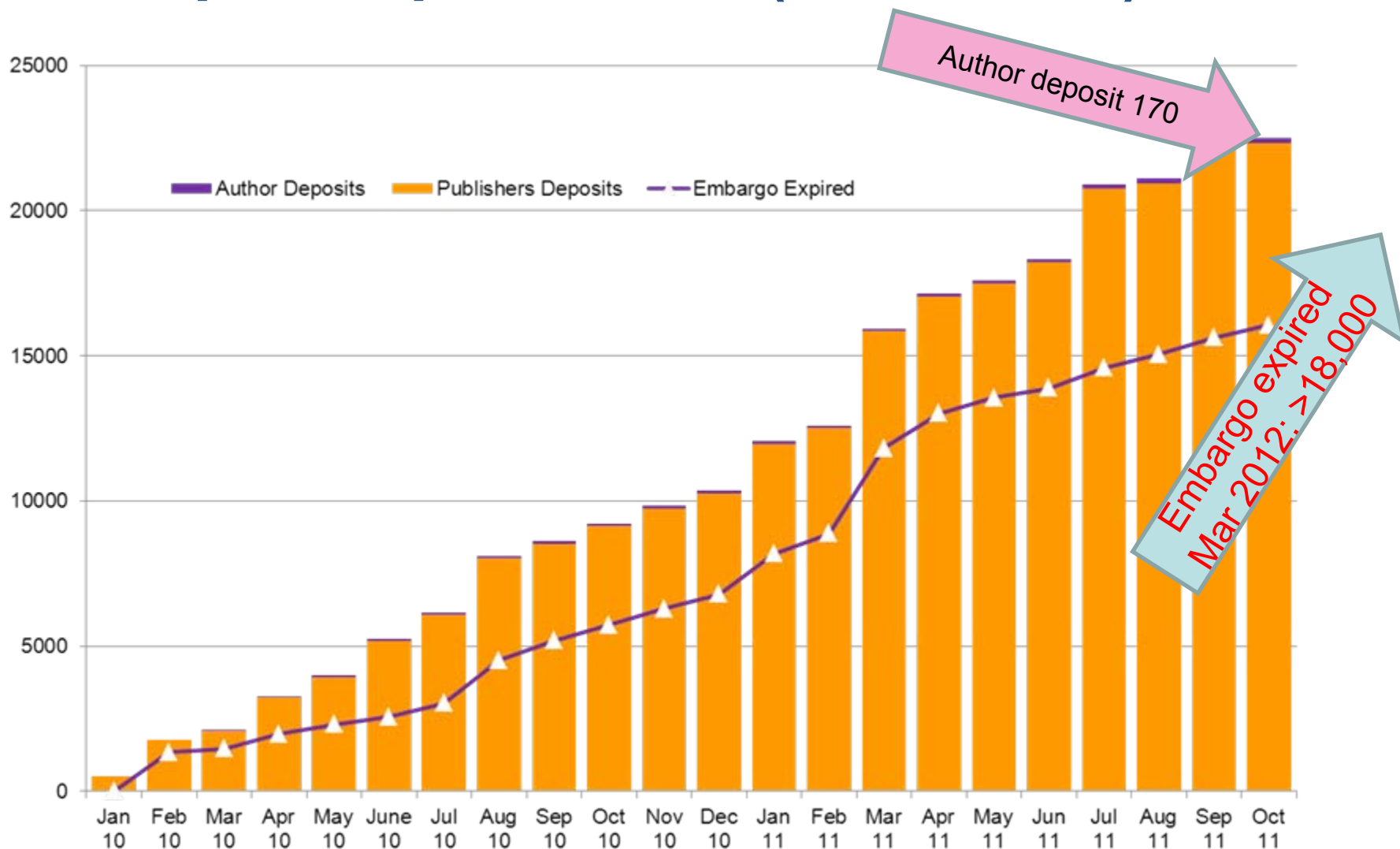
**“Observatory”**  
developed to monitor  
the impact of  
systematically  
depositing stage-two  
outputs on a large scale



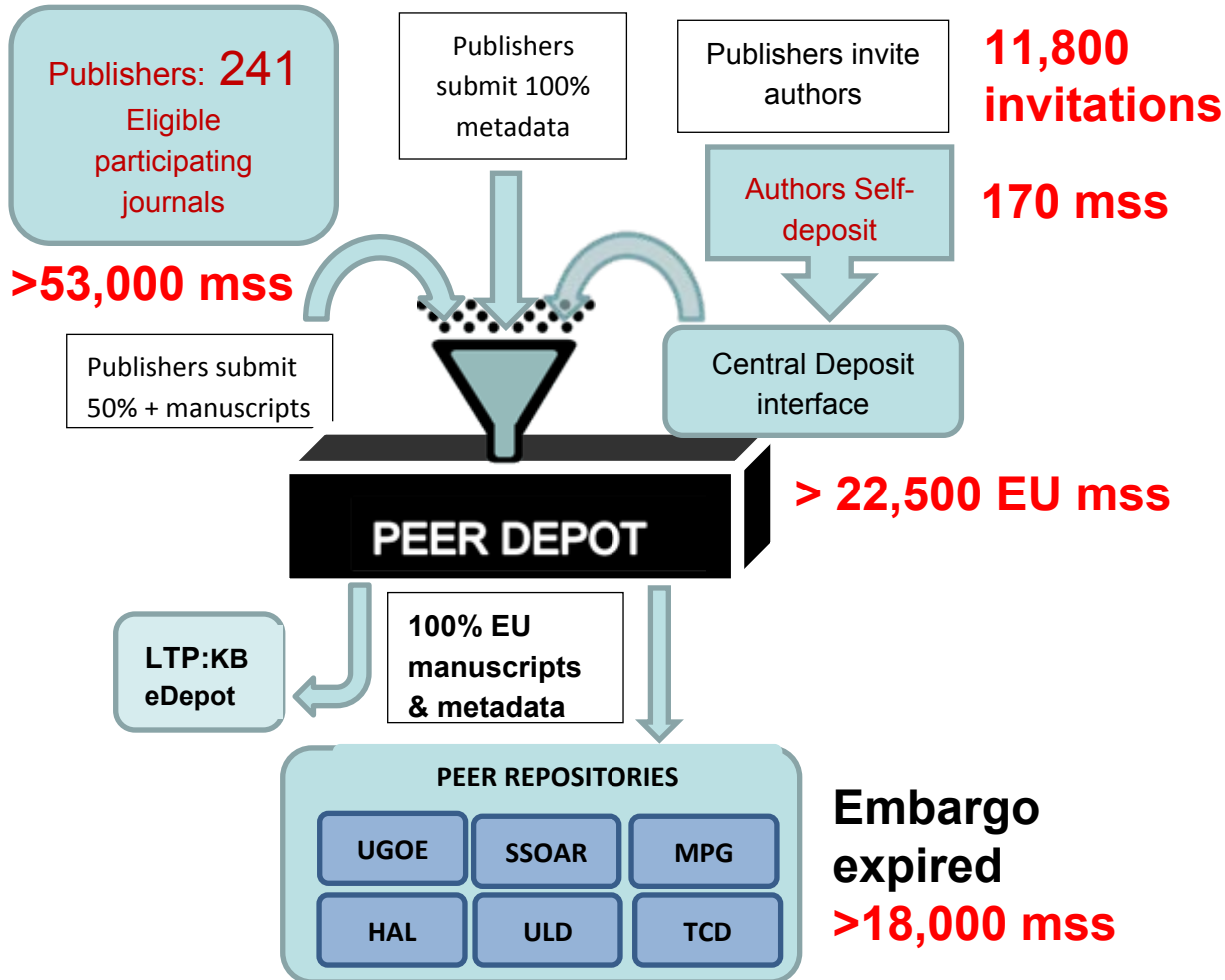
# Publisher deposits (cumulated)



# EU-Deposits processed (cumulated)



# 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

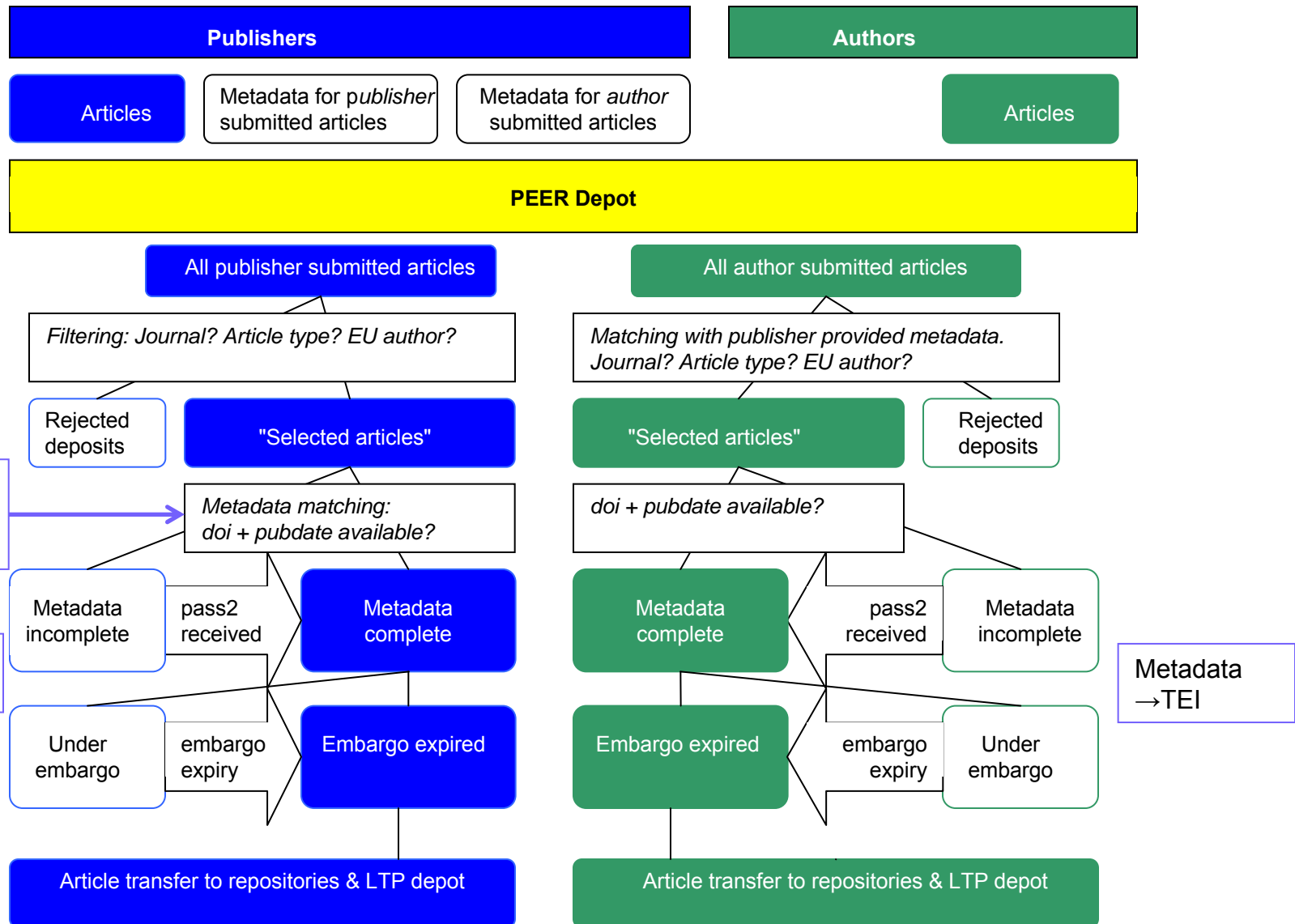
- Varying metadata requirements
- Varying ingestion processes
- Hosting PEER content
- Not configured for accurate embargo management
- Author authentication
- Logfile provision

## 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)



# Achievements to date

- **Enormous efforts made and results obtained**
  - Getting 6 heterogeneous repositories working in harmony on one project
  - Building the PEER Depot and creating infrastructural processes and protocols
  - Getting 12 very different publishers to contribute 241 test and over 200 control journals
  - Getting feeds for 241 heterogeneous journal systems to comply with PEER Depot requirements
  - Getting >53,000 mss processed the PEER Depot with uniform metadata
  - Ensuring that after EU filtering, each embargo group and subject has a statistically significant sample set of mss
  - Appointing and managing 3 leading research teams to work on the Observatory
  - **A working large-scale Observatory delivering results!**
- **Functioning collaborative infrastructure**
  - Linking repositories and publishers
  - Organising the transformation and flow of content
  - Metadata curation (quality control, embargo management etc.)
  - Usage data being collected from repositories and publishers
- **Substantial quantities of content visible in repositories:  
>18,000 EU deposits made publicly available (March 2012)**

# PEER Research Projects

<http://www.peerproject.eu/peer-research/>

- **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:

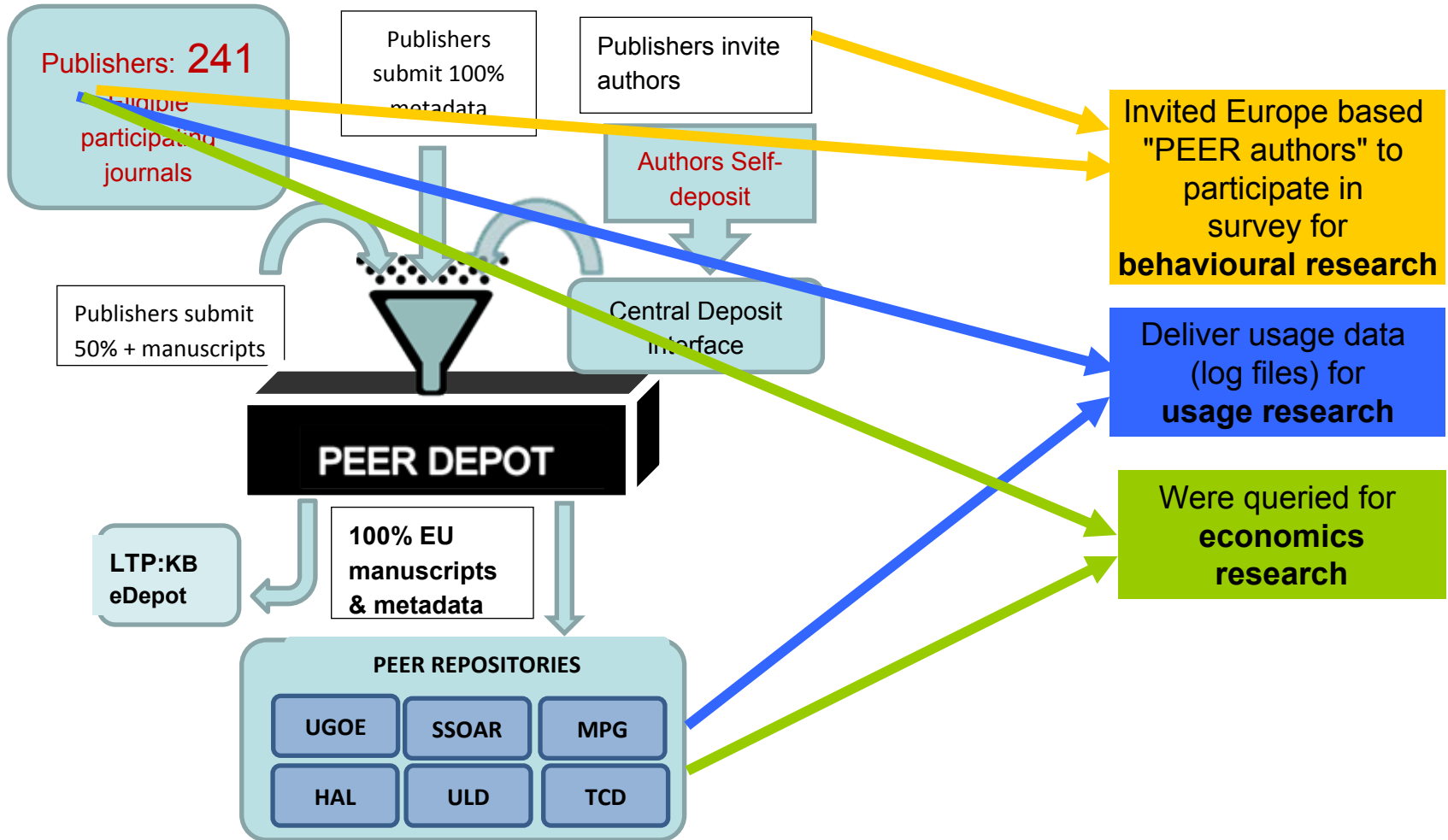
- Carol Tenopir, University of Tennessee (USA)
- Cherifa Boukacem, Lille University (France)
- Tomàs Baiget, *El profesional de la Información*, Barcelona (Spain)

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

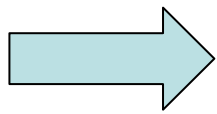
- Validate the specification for the research
- Advise on methodologies
- Evaluate the deliverables and confirm that the data is sound and conclusions are valid



# The PEER Observatory & Research



# Behavioural research



Unique collaboration with publishers, repositories and scientists to reach authors and users

Department of Information Science and LISU at Loughborough University, UK

## –Objectives

- Track trends and explain patterns of author and user behaviour in the context of so called Green Open Access.
  - Understand the role repositories play for authors in the context of journal publishing.
  - Understand the role repositories play for users in context of accessing journal articles.
- Two phases of Research between 2009 and 2011
    - adopted a mixed methods approach consisting of surveys, focus groups and an interdisciplinary workshop
    - Results covered by: **PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories. Final report (2011)** (available at <http://www.peerproject.eu/reports/>)

# Behavioural research: Final Report

## 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).
- Authors tend to be favourable to Open Access [...] 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.[...]
- Academic researchers [...] 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.
- There is anecdotal evidence that some researchers consider making journal articles accessible via Open Access to be beyond their remit.

# Economic research : Final Report

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



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

Article publication costs (earlier reports est. \$3-4k+):

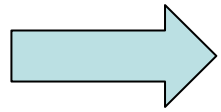
- Peer review has real costs and there are no economies of scale. (Average cost \$250 salary and fees only, excludes overheads - infrastructure, systems etc. and is heavily affected by rejection rates)
- Excluding peer review, average production cost ranges from \$170 to over \$400 per article (again excluding all overheads)
- Annual publisher platform maintenance costs ranges from \$170k to \$400k (excludes set up & development costs typically costing hundreds of thousands of dollars)

# Economic research: Final Report

- Repositories may have large sunk costs that are not accounted for
- Publishers (SB and OA) and repositories affected by *‘sustainability and competition for resources and reputation’*.  
Six key elements:
  - The growth in the supply of documents
  - The fragmentation of industry practices and interdependence among players
  - The incentives set up by a variety of research institutions
  - The economy of attention
  - The preservation of memory
  - The overall financial crisis and a generalised strain on resources

Results covered by: **PEER Economics Report** available at  
<http://www.peerproject.eu/reports/>

# Usage research: Logfile Analysis



First large-scale and comparative collection of article level usage

–CIBER Research Ltd., UK [<http://ciber-research.eu/>]

–Objectives:

- Determine usage trends at publishers and repositories;
- Understand source and nature of use of deposited manuscripts in repositories (so called Green Open Access) via usage data provision
- Track trends, develop indicators and explain patterns of usage.

# Usage research: Reports

- 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

**Results to be announced at  
PEER End of Project Conference  
May 29th 2012 Brussels**

Time: 11:00 - 17.00 CEST.

Federation of Enterprises in Belgium (FEB)

**Free registration at:**

**<http://www.peerproject.eu/peer-end-of-project-conference-29th-may-2012/>**