



**Dr Julia M Wallace**

**Project Manager – PEER**

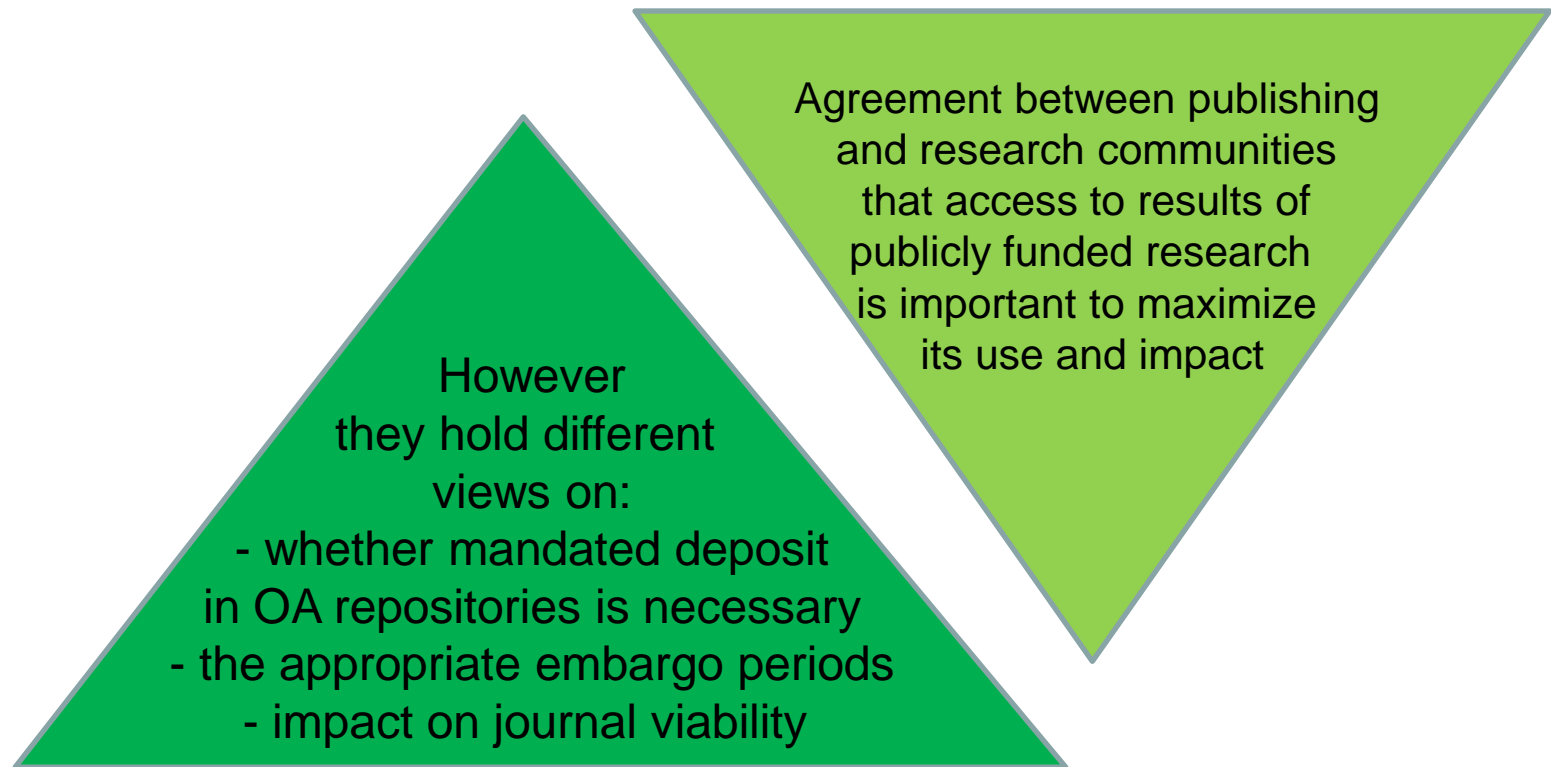
**Economy and acceptance of Open  
Access Strategies, 14 October 2011,  
Frankfurt**

# **The PEER Project: Green Open Access - experience, evidence and insights**

# Presentation overview

- PEER Project Overview
- The PEER Observatory
- Challenges & Solutions
- Achievements to date
- Research Studies in PEER

# Agreement and Disagreement



# Project objectives

- **PEER has been set up to investigate the effects of systematic archiving of 'stage two' research outputs (NISO: accepted manuscripts)**
- Large-scale 'experiment' regarding deposit of accepted 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

# PEER by numbers

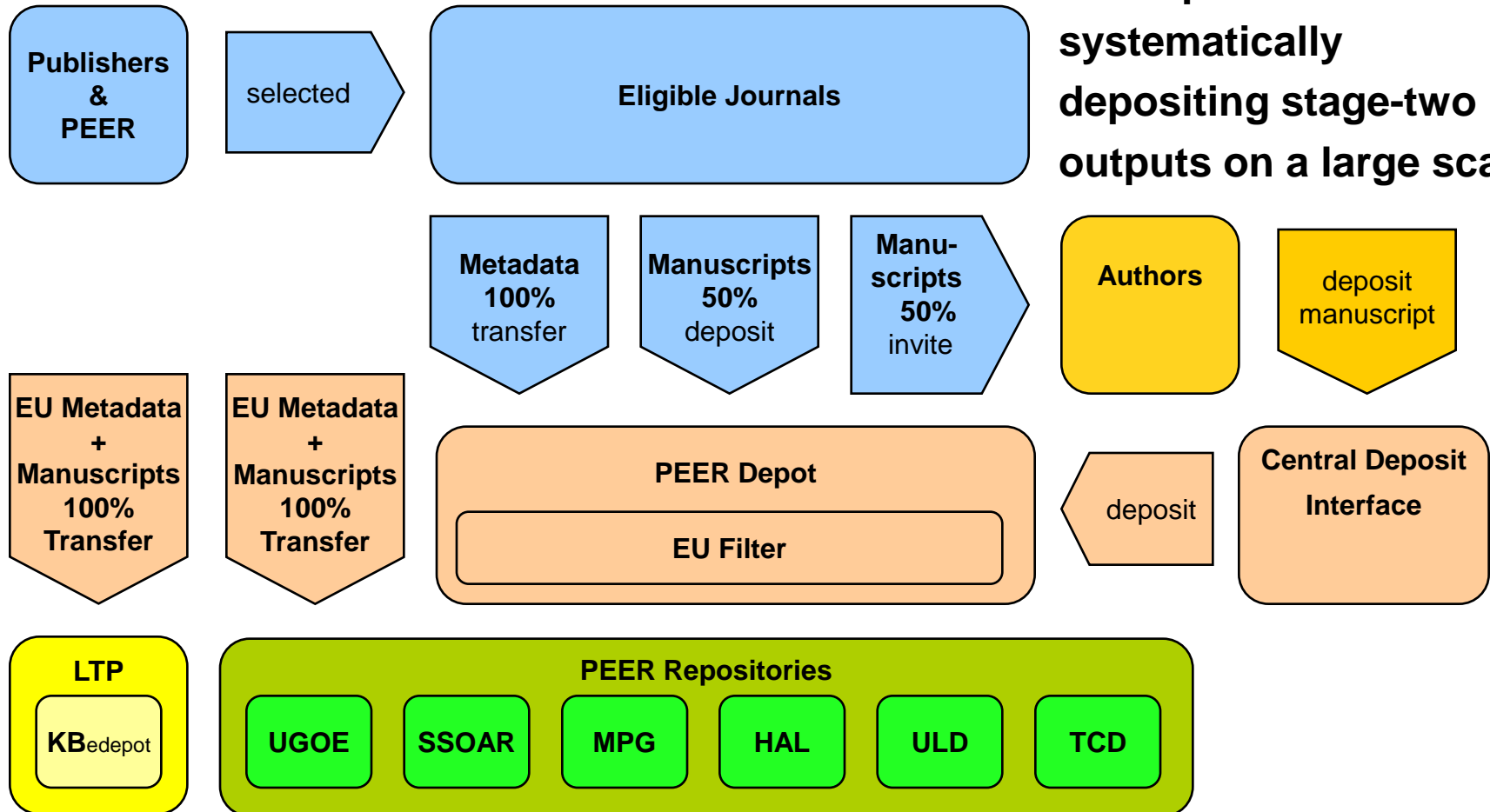
- Duration
  - 09/2008–05/2012 (3 years plus 9 months extension)
- Budget/Funding
  - €4.2M : 50% from the EC (eContent*plus* programme) 50% partners
- PEER participants
  - 5 Partners: STM (coordination), ESF, UGOE/SUB, MPG/MPDL, INRIA
  - 2 Technical partners: U. Bielefeld, SURF
  - 12 Publishers
  - 241 participating Journals (+>200 control journals)
  - 1 Depot/ Dark Archive
  - 6 Repositories
  - 1 Long-term preservation archive
  - 3 Research studies

# PEER Observatory

- The Observatory consists of
  - PEER Depot
  - PEER Repositories
  - plus usage data from the publisher platforms
- 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)
  - ~53,000 articles processed; ~16,000 EU deposits publicly available
  - 2 ways of articles deposit: publisher deposit / author self-archiving

# The PEER Observatory

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

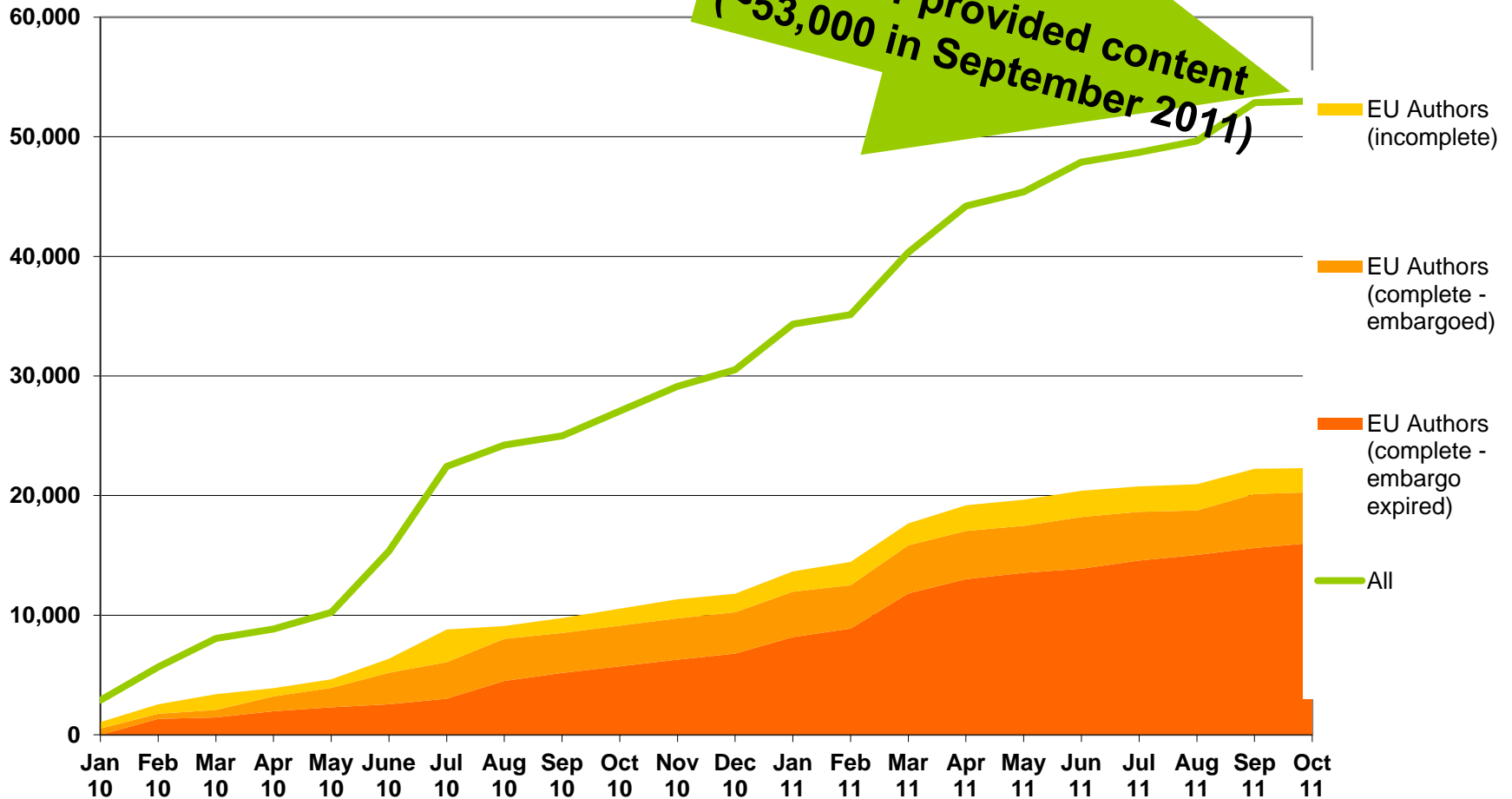


# Participating Publishers

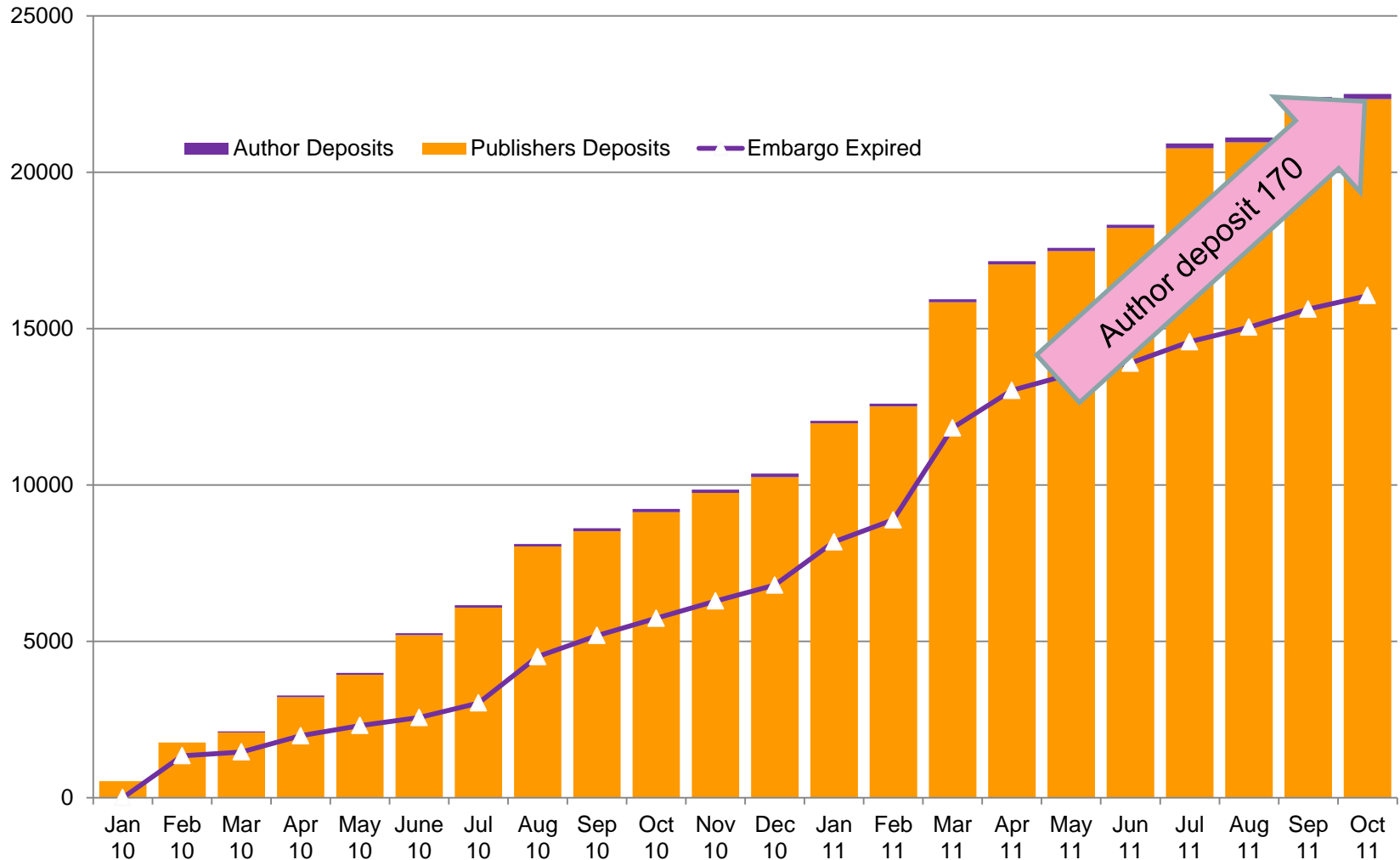
- BMJ Publishing Group
- Cambridge University Press
- EDP Sciences
- Elsevier
- IOP Publishing
- Nature Publishing Group
- Oxford University Press
- Portland Press
- Sage Publications
- Springer
- Taylor & Francis Group
- Wiley-Blackwell



# Publisher deposits (cumulated)



# EU-Deposits processed (cumulated)

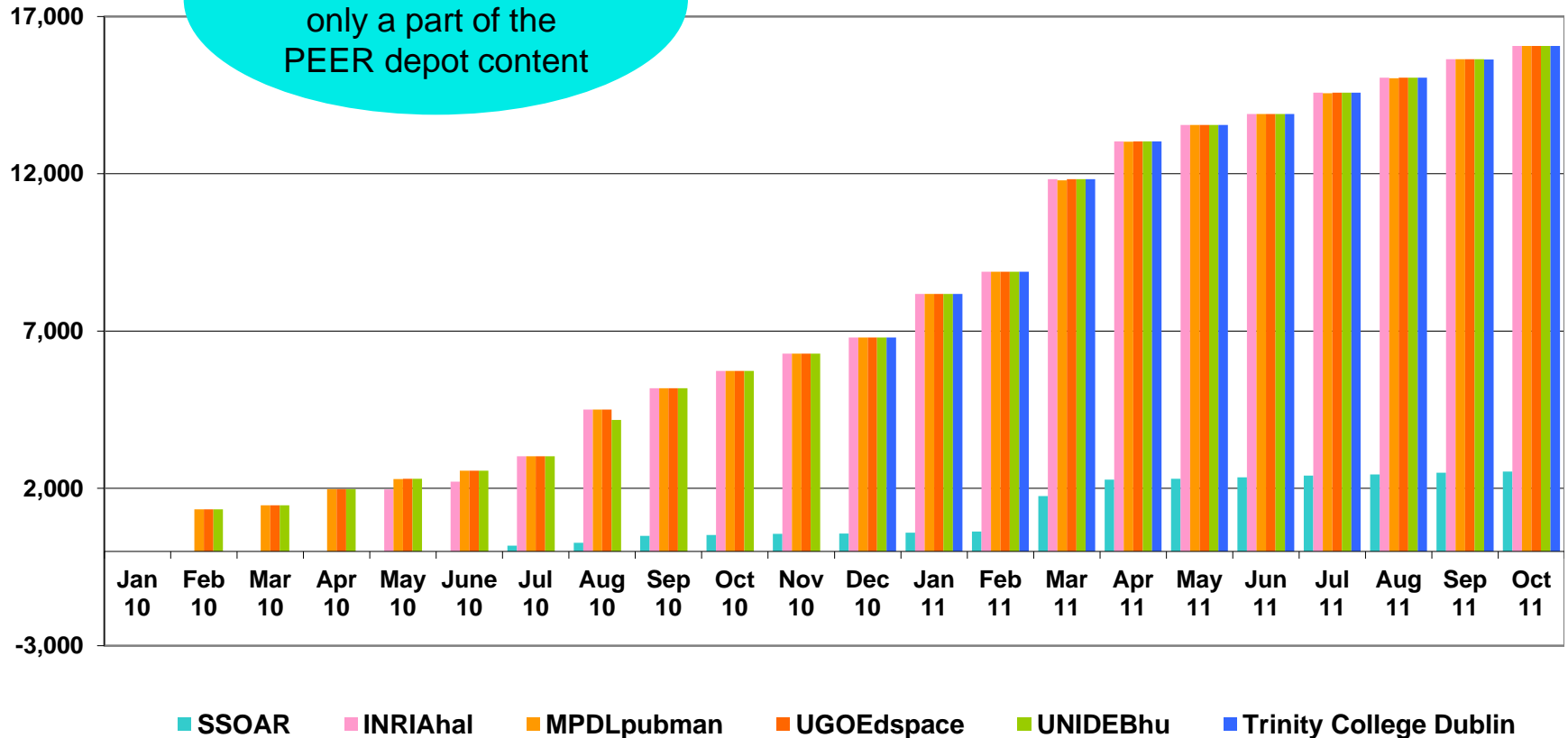


# Participating repositories

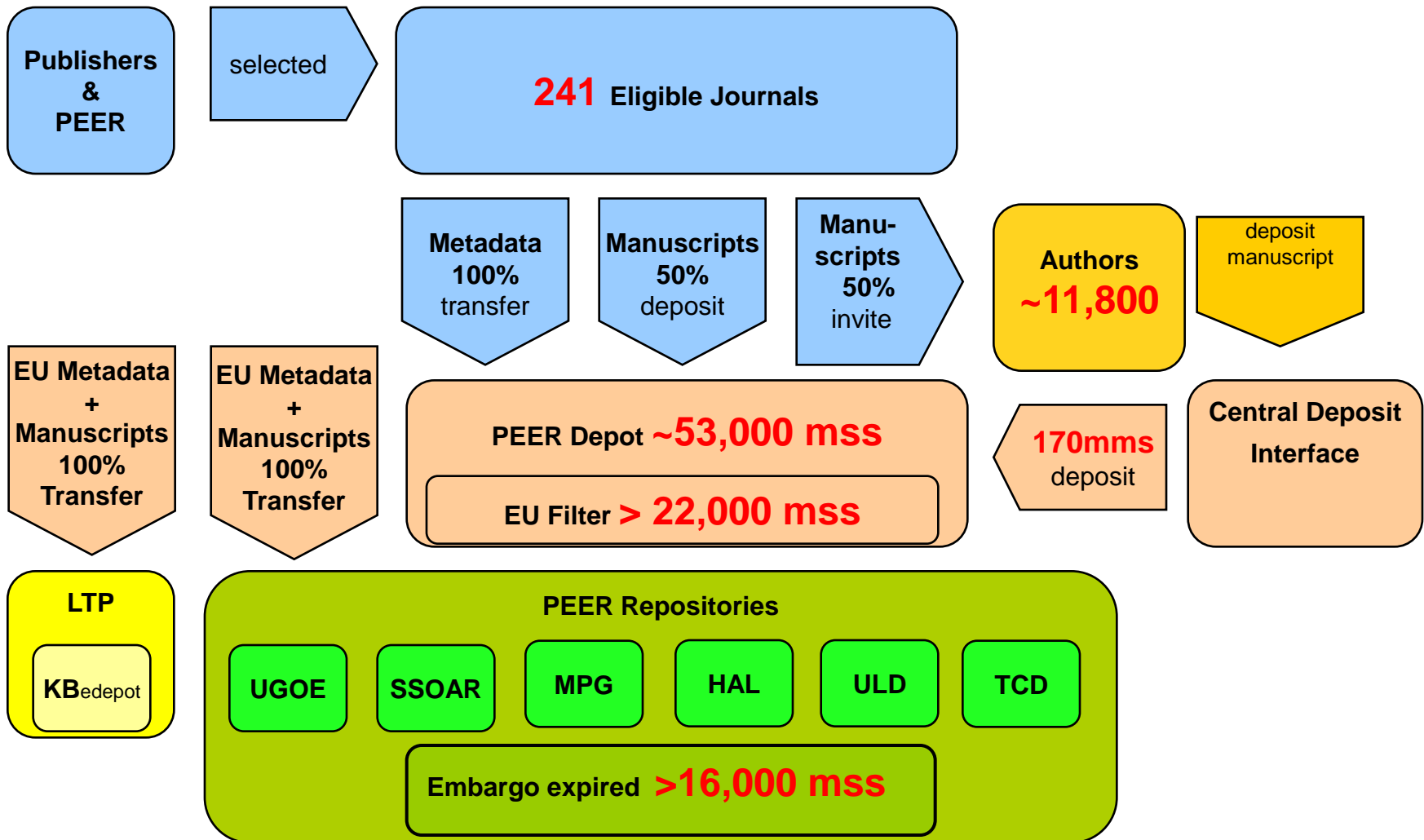
- eSciDoc.PubMan.PEER, Max Planck Digital Library (MPDL), Max-Planck-Gesellschaft zur Förderung der Wissenschaften e. V. (MPG)
- HAL, CNRS & Institut Nationalde Recherche en Informatique et en Automatique (INRIA)
- Göttingen State and University Library (UGOE)
- SSOAR – Social Sciences Open Access repository (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 repository population

SSOAR is a subject repository accepting only a part of the PEER depot content



# The PEER Observatory



# Challenges and Solutions – Publishers

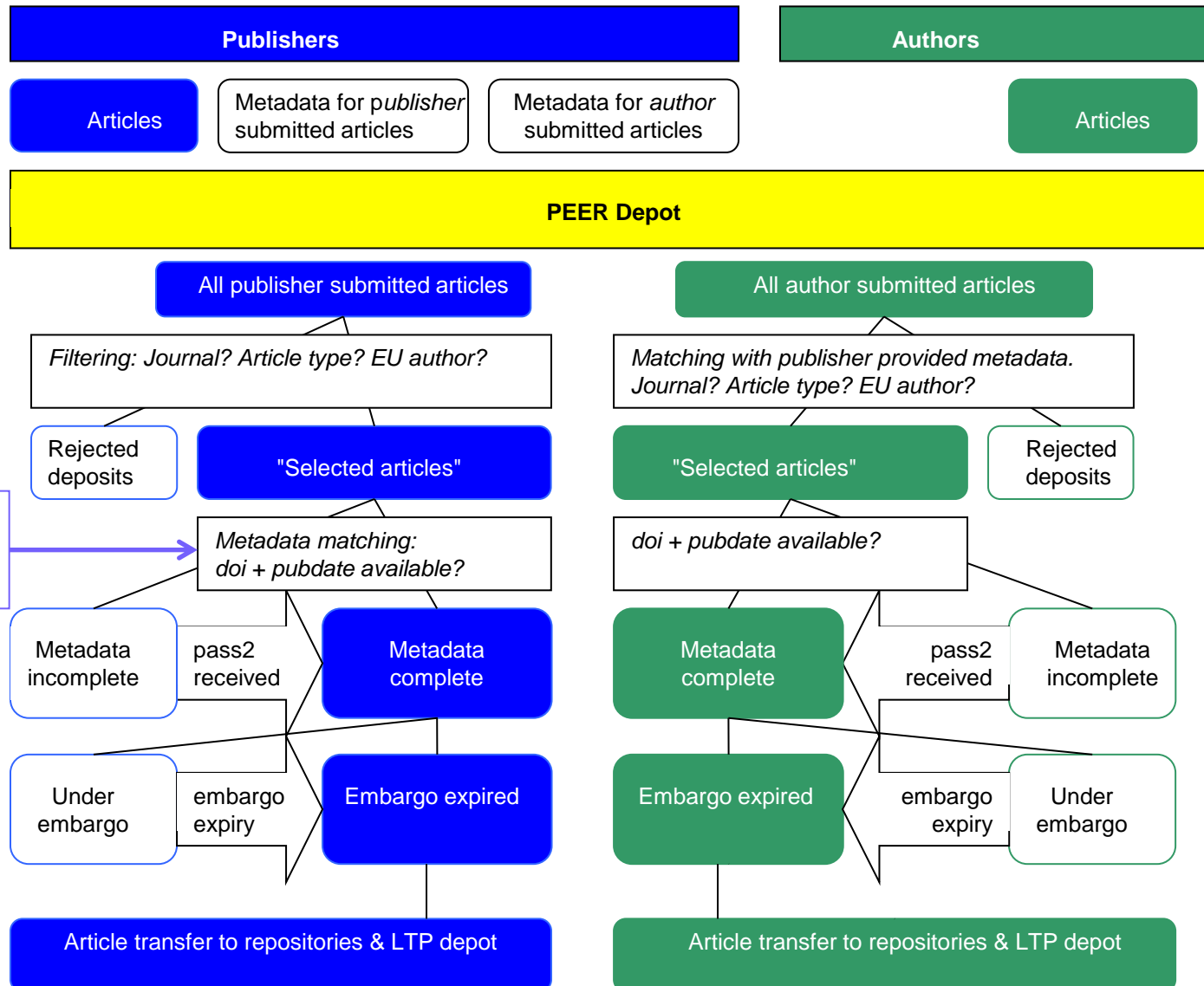
- **Ensuring journal validity**
  - Correct article type and matching metadata
  - All mandatory metadata (publication date!) received
- **EU corresponding author**
- **Metadata delivery in several batches**
  - Article metadata are incomplete at acceptance time; Publication date unknown, DOI not attributed
  - Extraction of only „EU“ authored manuscripts not possible at acceptance stage
- **Author accepted manuscripts in a variety of file formats**
- 
- **Different metadata formats**
  - NLM2.x, NLM 3.0, ScholarOne, proprietary
- **Metadata delivered within PDF document**
- **Non publisher related technical challenges** (Author authentication, embargo management, file formats/metadata required by repositories)

- **Checking mechanisms**
  - Document kept until metadata completion
- **Author affiliation check/ filter**
- **Article kept until metadata completion**
  - Metadata are accepted in either one step (on publication) or two passes (on acceptance and on publication)
  - Extraction done at PEER Depot
- **Only one file format allowed – PDF**
- **Mapped into single TEI structure**
- **Extraction done at PEER Depot (GroBID) in order to increase content**
- **Changes, adjustments at publishers**

# Challenges and Solutions – PEER Depot

- Deposit channels established for 12 publishers
- Link with PEER Author submission interface & match with publisher submitted metadata
- Validate file & metadata integrity: follow-up on problem areas
- Filter for valid participating journals (title, ISSN)
- Filter for EU authored content
- Filter for article type ‘valid research articles’
- Map publisher metadata schemas to PEER schema
- Transform metadata via TEI customisation
- Identify publication date & manage embargo period
- Distribute to participating repositories (SWORD protocol) & LTP archive (FTP)
- Publisher exception: Extraction of metadata from PDFs (via Grobid (*GeneRation Of Bibliographic Data*) – an automatic metadata extraction from PDF process
  - **This is a ground-breaking technical development**
- Repository exception: Filtering content on subject via journal sub-set (subject repository)

# PEER Depot Workflow





# Challenges and Solutions – Repositories

- **Technicalities/ Arrangement of Data acceptance**
  - Adjusting to PEER Standard
  - Implementation of SWORD protocol
  - Build dedicated PEER Repository within framework of home institution
  - Convert TEI metadata into internally used metadata standard
  - Anonymisation of log files
  - Set up automated log file transfer to Usage Research Team
  - To be responsive regarding current issues
- **Difficulties encountered when inviting to join the Repository Task Force**
  - Thematic focus of manuscripts dealt with in PEER does not meet the bias of the invited repository
  - Find it difficult to comply with PEER standards
  - Find it difficult to communicate with the project (resources)
  - Find it difficult to make resources available (financial/ personnel / technical)

## Green OA environment issues encountered by PEER

- Non uniformity of publisher outputs
- Varying requirements by repositories
- EU & article type filtering of content
- Embargo management
- Author authentication for deposit
- Non uniformity of log files
- Format problems with back-content files
- Technical & financial challenges for repository participation (non PEER Partner repositories)

# Achievements to date

- **Enormous efforts made and results obtained**
  - 6 heterogeneous repositories working in harmony on one project
  - Building the PEER Depot and creating infrastructural processes and protocols
  - 12 very different publishers contributing 241 test and over 200 control journals
  - Ensuring feeds for 241 heterogeneous journal systems to comply with PEER Depot requirements
  - ~53,000 mss processed by the PEER Depot with uniform metadata
  - 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:  
~16,000 EU deposits made publicly available (as of September 2011)**

# PEER Research Projects

- **Commissioned from independent research teams**
  - **Usage research** (CIBER Research Ltd. – David Nicholas and Ian Rowlands)
  - **Economic research** (ASK Bocconi – Paola Dubini)
  - **Behavioural research** (LISU and DIS, Loughborough – Jenny Fry and Claire Creaser)

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

# PEER Research Projects

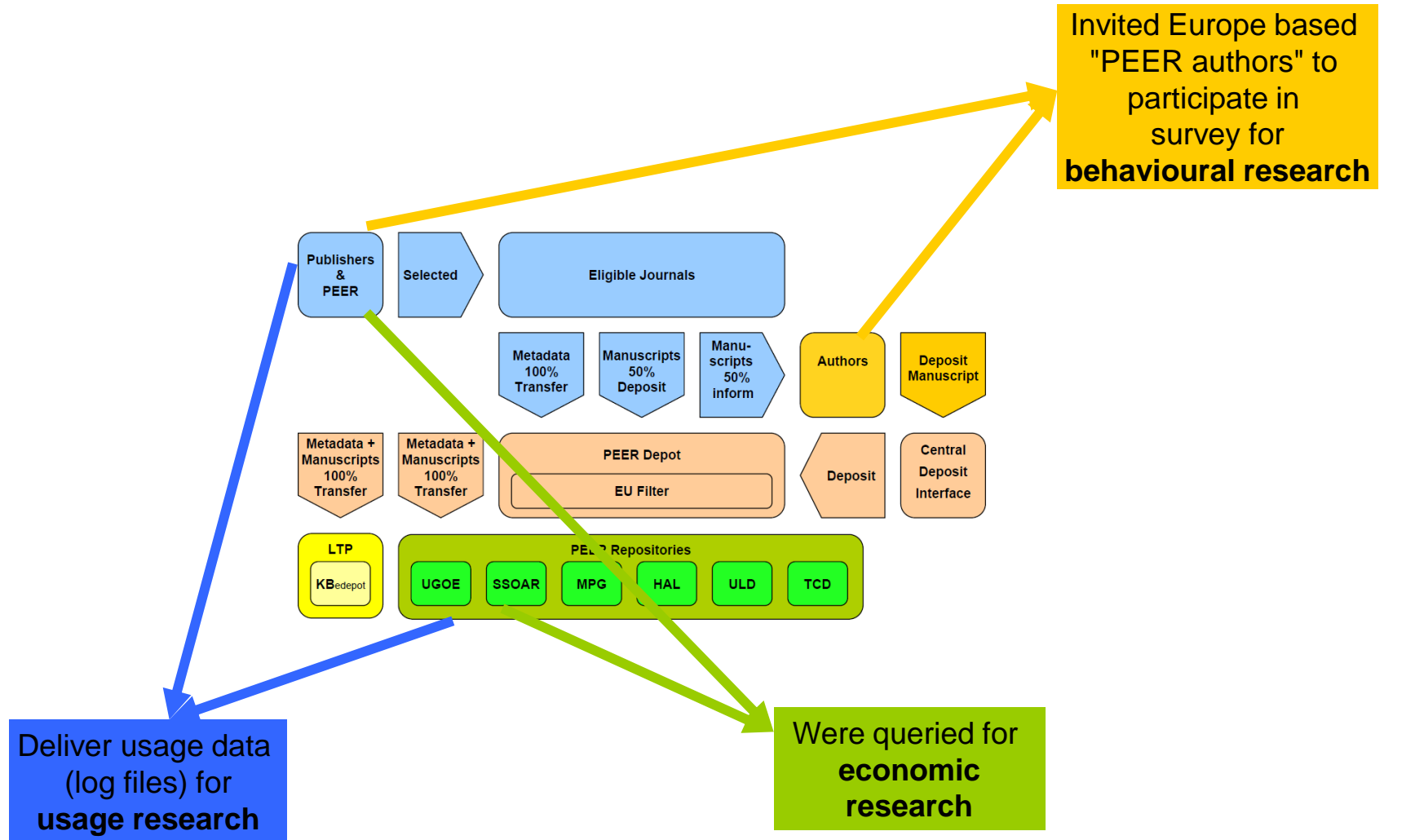
- High-quality, credible research, neutral, transparent and supported by all stakeholder groups

- **Research Oversight Group (ROG)**

Expert panel comprising three independent experts in scholarly communications and publishing:

- Carol Tenopir, University of Tennessee (USA)
  - Cherifa Boukacem, Lille University (France)
  - Tomàs Baiget, *El profesional de la Información*, Barcelona (Spain)
- 
- Validate the specification for the research
  - Advise on methodologies
  - Evaluate the deliverables and confirm that the data is sound and conclusions are valid

# Peer Observatory + Research Projects



# Usage research: Logfile Analysis

– CIBER Research Ltd., UK

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.



**First large-scale and comparative collection of article level usage**

- At Publishers: 241 participating journals (items with EU authors), 200 control journals
- At repositories: embargo-expired items

# Economic research

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

## **Objectives**

- Analyze the overall effects of large-scale deposit (Green OA) on the economics of scholarly communication.
- Investigate the cost of the large-scale deposit of stage-2 research outputs; including the economic efficiency or cost of the process of deposit.
- Understand the costs incurred by participating publishers and PEER repositories
- Understand, principally, for the deposit of so-called Stage 2 manuscripts the costs a) in time to depositors; b) for the set-up and the longer term to repositories and/or libraries; and c) to publisher when co-operating in the deposit process



**First detailed empirical study of cost drivers of publishers and repositories – 22 case studies**



# Behavioural research

**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



**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

## KEY CONCLUSIONS (1)

- Over the period of Phases 1 and 2 of the Behavioural research project the increase in the number of researchers who reported placing a version of their journal article(s) into an Open Access Repository was negligible.
- 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 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.

# Behavioural research: Final Report

## KEY CONCLUSIONS (2)

- There is anecdotal evidence that some researchers consider making journal articles accessible via Open Access to be beyond their remit.
  - 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.
- ... **Read more at:** <http://www.peerproject.eu/reports/>

# Announcement: Final PEER CONFERENCE: May 29th 2012 Brussels



Details will be announced on the PEER website: <http://www.peerproject.eu>

**Recent Article:** PEER, green open access - insight and evidence; *Learned Publishing*, 24:267–277 doi:10.1087/20110404

**Thank you  
for your attention!**

**Questions?**

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

**or**

**e-mail: [peer@stm-assoc.org](mailto:peer@stm-assoc.org)**