



# PEER Project: Challenges & Solutions

Open-Access-Tage 2011



Barbara Kalumenos

Niedersächsische Staats- und  
Universitätsbibliothek Göttingen

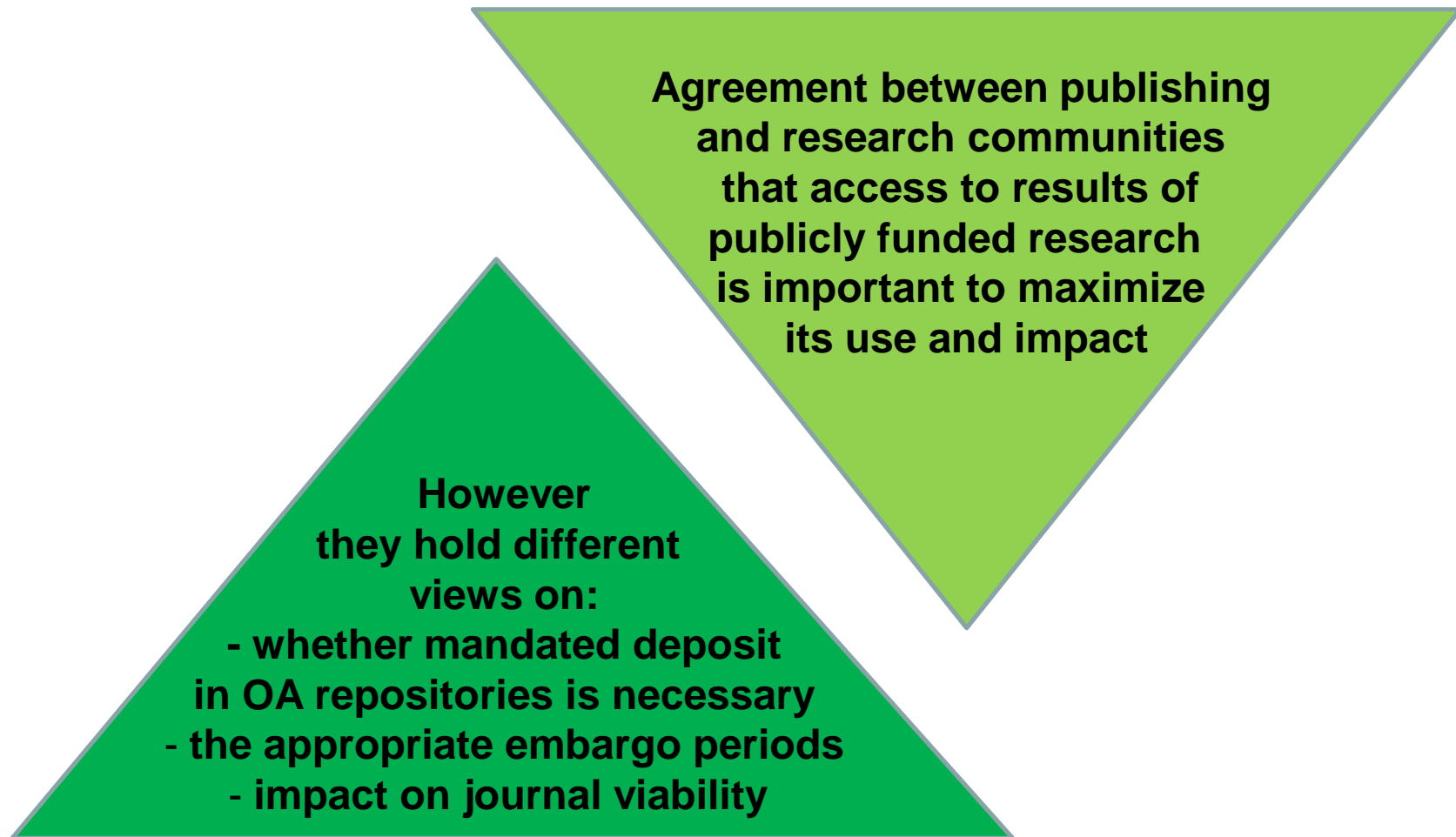


Barbara Bayer-Schur

# Agenda

- Overview
- The PEER Observatory
- Challenges & Solutions - Publishers
- Challenges & Solutions - PEER Depot
- Challenges & Solutions - Repositories
- Achievements to date
- Research Studies in PEER

# Agreement and Disagreement



# Project Overview

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

# Project objectives

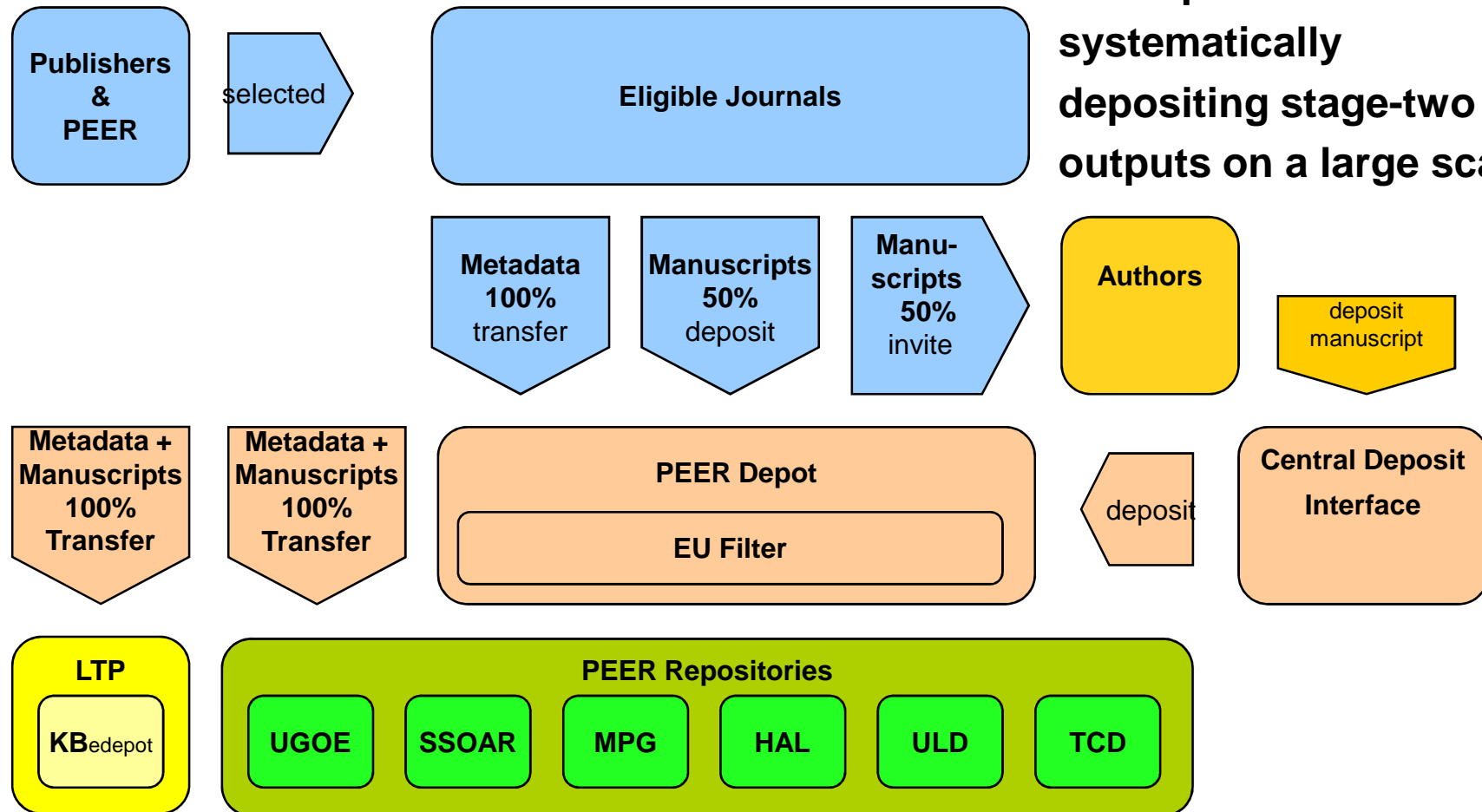
- **PEER has been set up to monitor the effects of systematic archiving of 'stage two' research outputs**
- Large-scale ,experiment' regarding deposit of author manuscripts: in an 'observatory' of OA repositories
- Research studies to gather hard evidence to inform future policies
  - Usage Research → Availability, usage
  - Behavioural Research → Author behaviour
  - Economic Research → Costs, viability
- Collaborative project of diverse stakeholder groups
  - Publishers, research community and library/repository community
  - **“Report on economic research comparing publishers-assisted deposit with self-archiving”**

# PEER Observatory

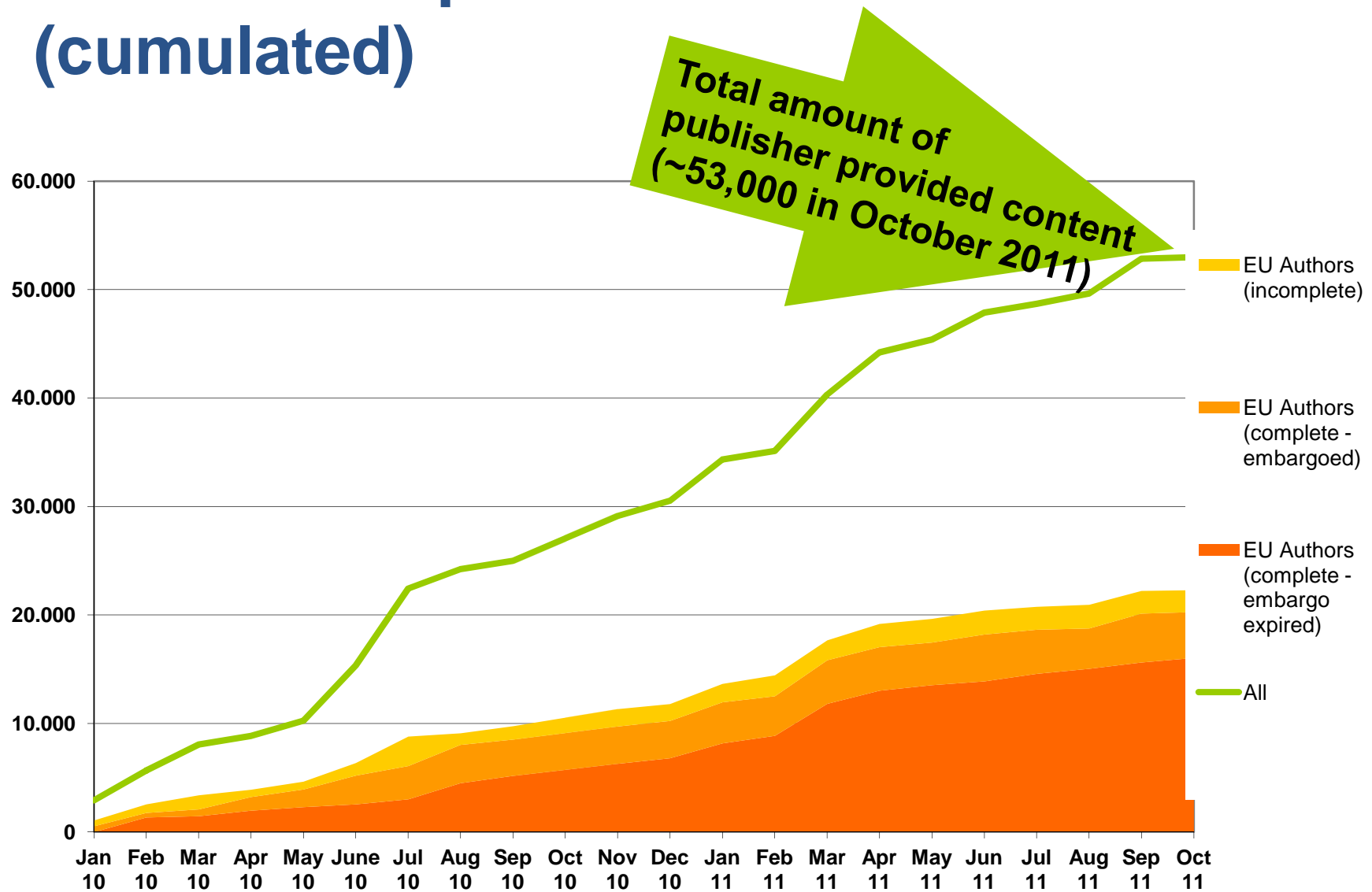
- The Observatory consists of the
  - PEER Depot
  - PEER Repositories
- The Depot
  - Acts as a „Clearing House“ - is a Dark Archive!
  - Processes deposits
  - Distributes content to participating repositories
- The PEER Repositories
  - Provide the usage data (= log files) needed by our research partner CIBER
- Content inflow
  - 241 journals from four broad areas; selecting process
  - ~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

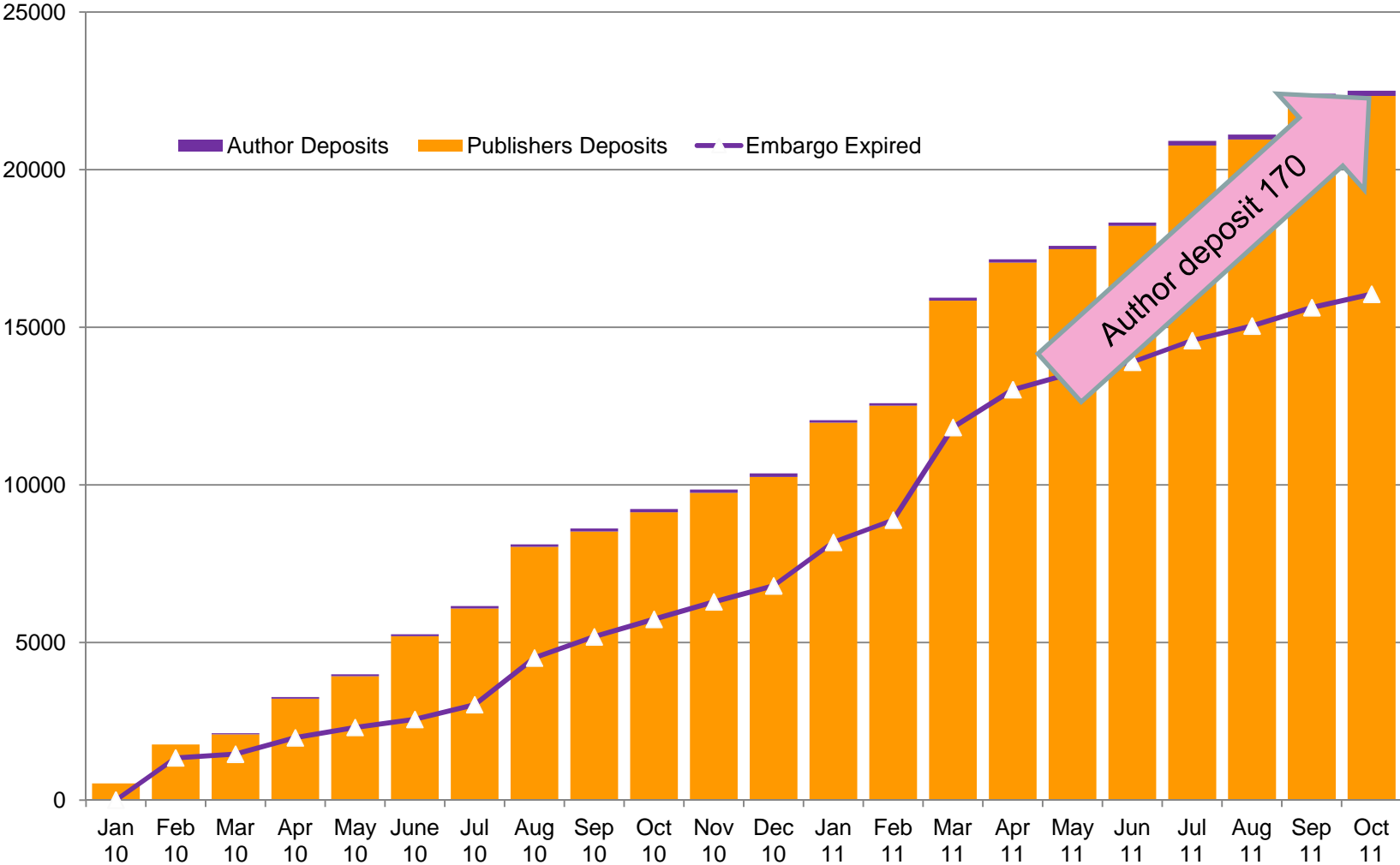


# Publisher deposits (cumulated)



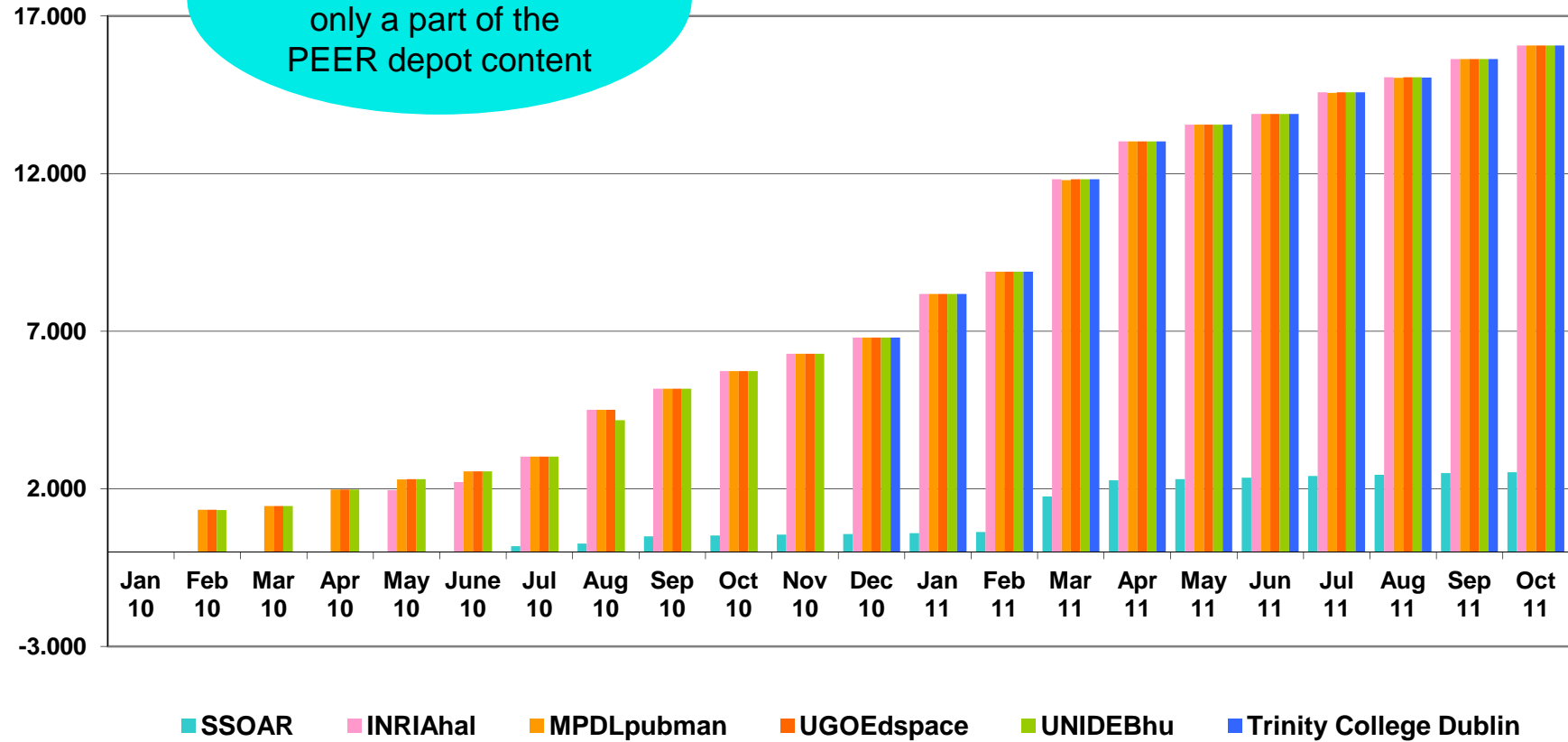


# EU-Deposits processed (cumulated)



# PEER repository population

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



# Challenges and Solutions – Publishers

- **Ensuring journal validity**
  - Correct article type and matching metadata
  - All mandatory metadata (publication date!) received
- **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
- **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**

- **Agreement of Data transfer** (Face to face meetings; teleconferences; publisher specific communications)
- **STM – Publisher Cooperation** (Quartely meetings; progress reports; request for additional content)
- **Technicalities** (Some publishers amended their workflows)

# 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)

# 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 DC metadata internally
  - Anonymisation of log files
  - Set up automated log file transfer to Usage Research Team
  - To be responsive regarding current issues
- **Communication**
  - E-Mail communication and work package/ task force meetings
  - Face-to-face meetings & collective teleconferences
  - Repository specific communication with WP leader and Depot
- **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 to PEER standards
  - Find it difficult to communicate within the project
  - Find it difficult to make resources available

## 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
  - Lower than anticipated EU% of research content
  - Format problems with back-content files
  - Technical & financial challenges for repository participation (non PEER Partner repositories)
- *Delays within the project are due to innovation & change, technical and other challenges which could not have been foreseen at earlier stages of the project and are simply due to its complexity.*

# 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:  
~16,000 EU deposits made publicly available (as of September 2011)**

# PEER Research Projects

- Usage research
- Economic research
- Behavioural research

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



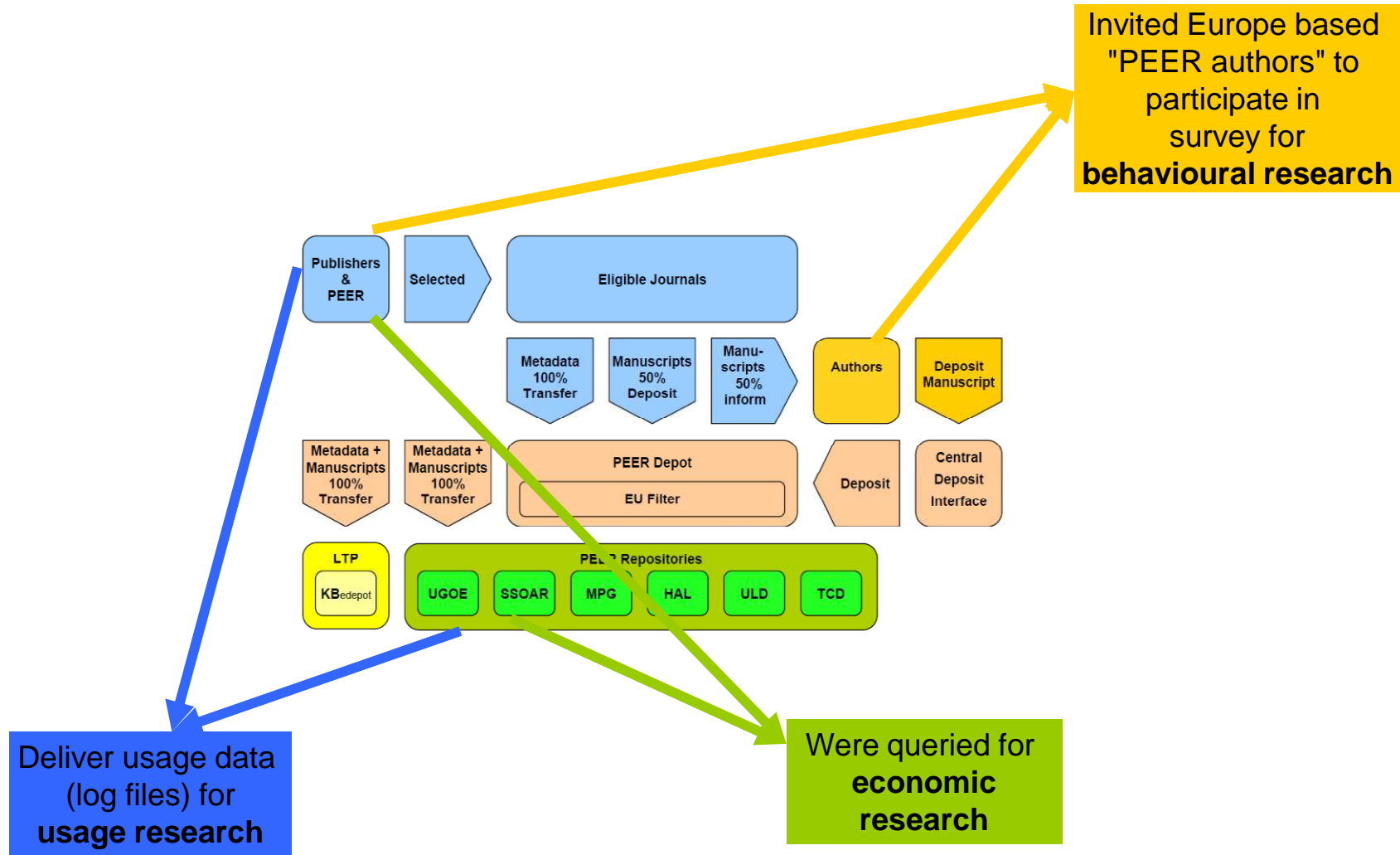
## PEER Research Projects

- High-quality, credible research, neutral, transparent and supported by all stakeholder groups
- To ensure this, the project needed to put up with delays, therefore 9 months extension
- **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)
  - 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 [<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.  
A contribution to the new field of usage research is expected.



First large-scale and comparative collection of article level usage

# Usage research: Final Report

- High volume of content in the project:  
**16,000 EU deposits** made publicly available  
CIBER requested sample of this size available for research with a high degree of confidence
- Ongoing: measure activity over 12 months, starting March 2011
- Log file collection until 31 August 2011 (first cut-off point for analysis)
- Interim confidential reporting: Sept 2011 (after 6 months),  
Dec 2011 (after 9 months), Feb 2012 (end)
  
- **“Final Report on usage research“**: April 2012  
available via <http://www.peerproject.eu/reports/>

# 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 to publishers and repositories

# Economic research: Final Report

→ “Report on economic research comparing publishers-assisted deposit with self-archiving“

## From the contents...

- Publishers cost structure
  - Cost of content certification
  - Cost of content publication
  - Cost of content archival and preservation
  - Compliance to PEER
- Repository cost structure
  - Cost of content uploading
  - Repository set up and maintenance cost
  - Author involvement
  - Involvement in PEER
- Comparison of cost structure and cost drivers
- The ecology of scholarly publishing
- Limitations of the study and recommendation

... Read more soon: <http://www.peerproject.eu/reports/>

# 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
  - Phase 1 (2009): Extensive Survey of European researchers + Focus groups
  - Results covered by: **PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories. Baseline report** (2010) (available at <http://www.peerproject.eu/reports/>)



Unique collaboration with publishers and scientists to reach authors and users

# Behavioural research: Final Report

→ “PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories. **Final Report**”

- Covers Phase 2 of the Research: 2nd Survey of European researchers + Final Workshop (2010-11), and provides a synthesis of the results of both phases

## From the contents...

- Awareness of Open Access and Open Access Repositories
  - Different types of repositories
  - Disciplinary similarities and differences
- Open Access Repositories in the research process
  - The use of Open Access Repositories by readers
  - Authors' open access behaviour and self-archiving practice
- Researchers' perceptions of the best way(s) of achieving Open Access
  - Motivations to self-archive and perceived benefits of OA
  - Perceived barriers of Open Access Repositories
  - Perceived influence of OAR on scholarly publishing
  - Disciplinary similarities and differences

... *Read more soon:* <http://www.peerproject.eu/reports/>



## **Announcements:**

- **Article: PEER, green open access - insight and evidence; Learned Publishing, 24:267–277  
doi:10.1087/20110404**
- **Final PEER CONFERENCE: May 29th 2012 Brussels**

**For details please visit our Webpage: <http://www.peerproject.eu>**

**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)**