



# The PEER Project: Observing the Impact of Green Open Access

**CERN workshop on Innovations in  
Scholarly Communication (OIA7)**



**Barbara Kalumenos**



**MAX PLANCK**  
**digital library**

**Christoph Bruch**

**The PEER project is  
about the green road  
to Open Access,  
also called self-  
archiving.**

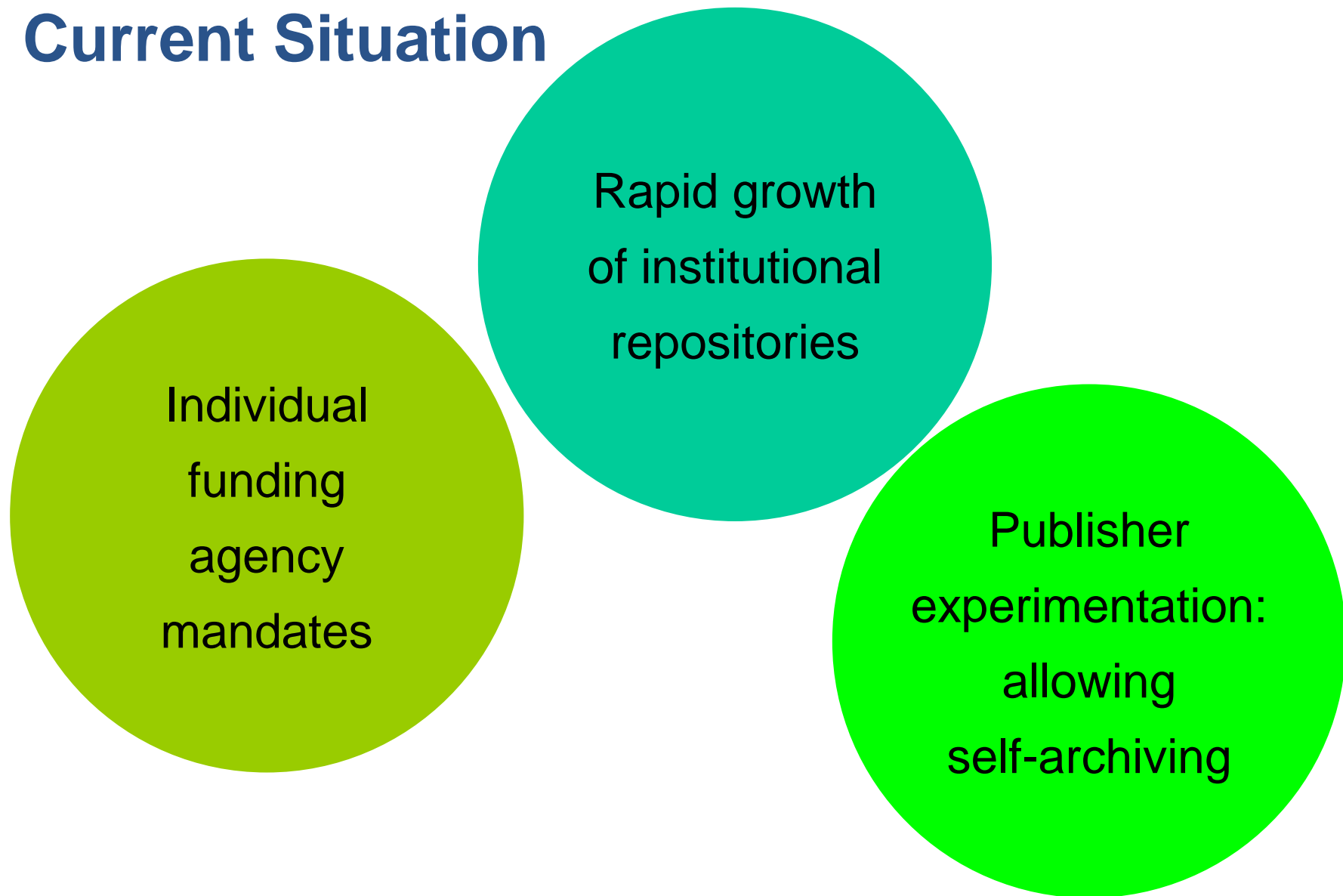
# Agenda

- Overview (4-8)
- The PEER Observatory (9-14)
- Challenges and Solutions - Publishers (15-18)
- Research Projects (19-25)

# Project Overview

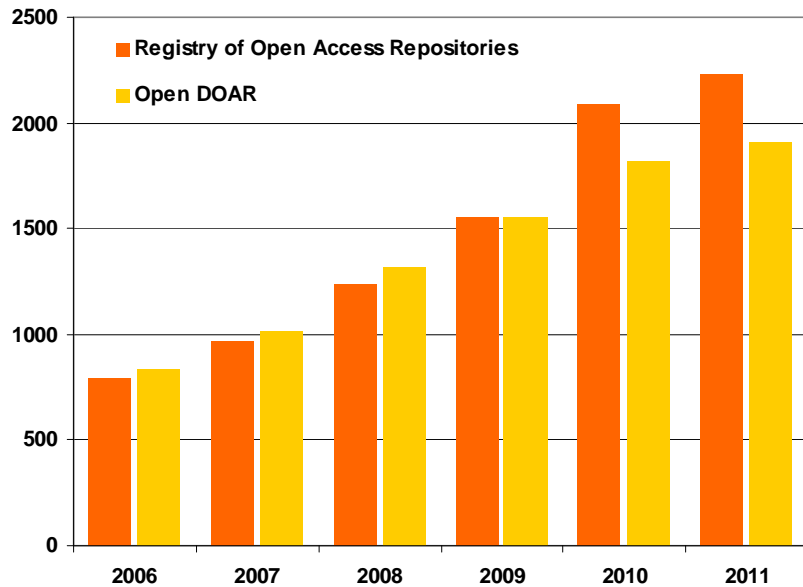
- Duration
  - 09/2008–05/2012 (3 years plus 9 months extension)
- Budget/Funding
  - 4.2 Mio €; 50 % by the European Union (eContentplus programme)
- Project partners
  - STM (coordination), ESF, UGOE, MPG/MPDL, INRIA
  - Technical partners: U. Bielefeld, SURF, KB Netherlands (long-term archiving)
  - 12 publishers
  - 6 repositories
- Contact / Website
  - [peer@stm-assoc.org](mailto:peer@stm-assoc.org) / <http://www.peerproject.eu>

# Current Situation

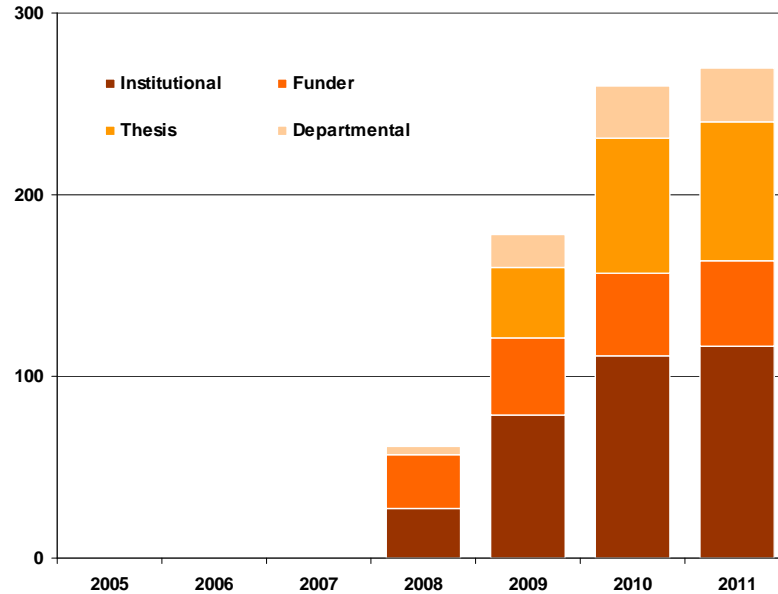


# Current Situation

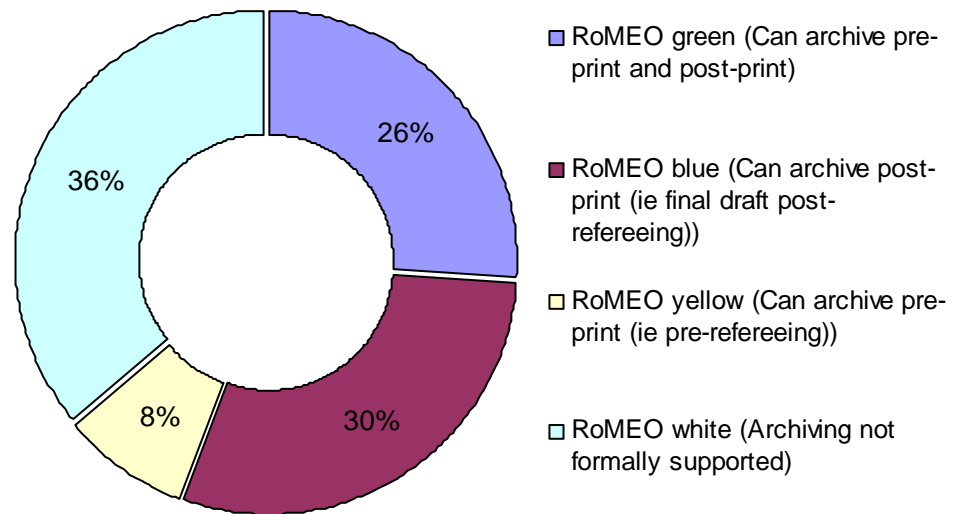
Number of Open Access Repositories



Open Access Mandates / Policies (ROARMAP)



Publisher's allowing green Open Access policies (RoMEO)



# Agreement and Disagreement

**There is agreement between publishing and research communities that access to the results of publicly funded research is important to maximize its use and impact.**

**However,**

**they hold different views on**

- whether mandated deposit in open access repositories is necessary**
- the embargo periods that would be appropriate.**
- the impact this may have on journals**

# 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 group
  - Publishers, research community and library/repository community

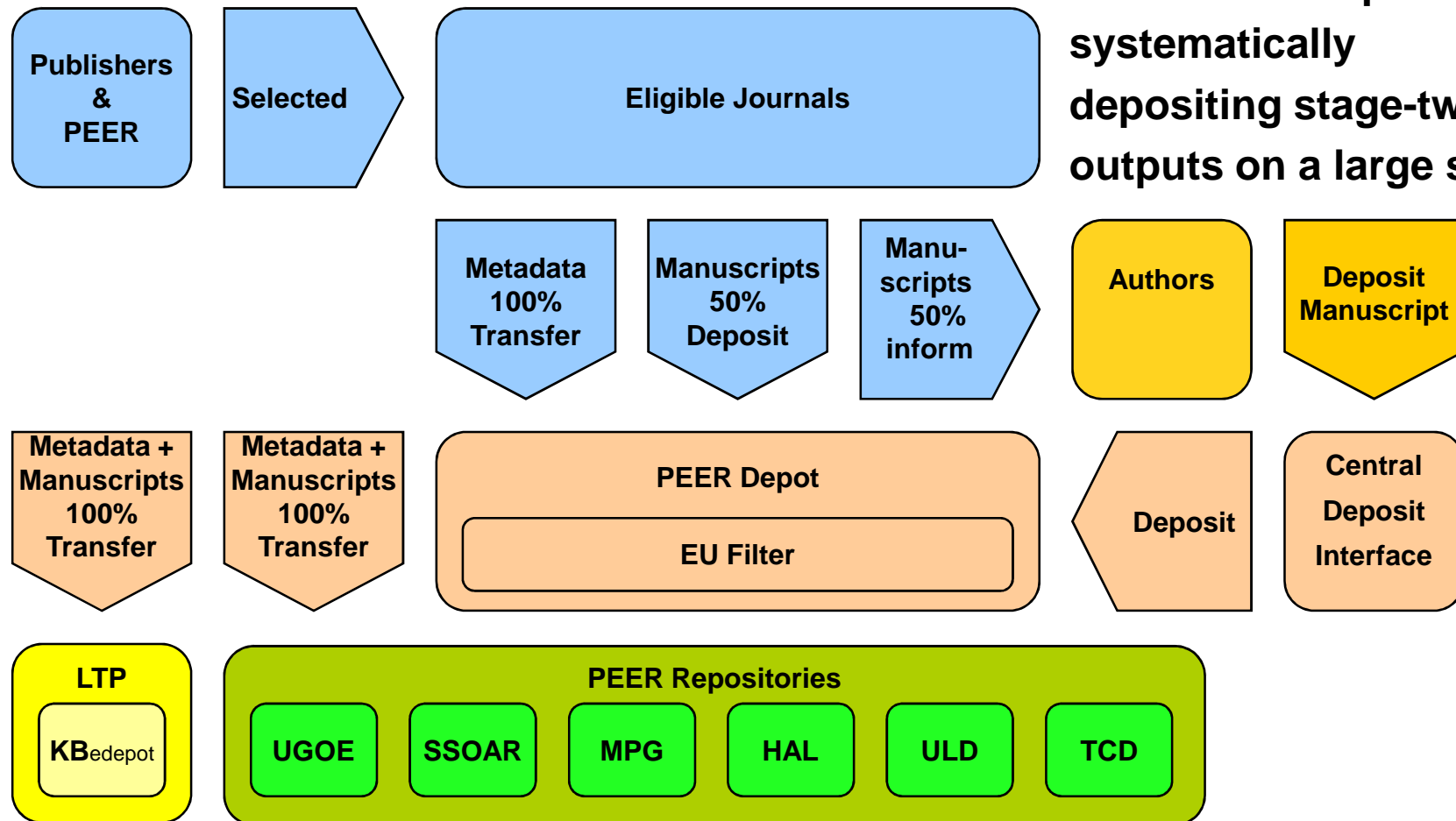


# PEER Observatory

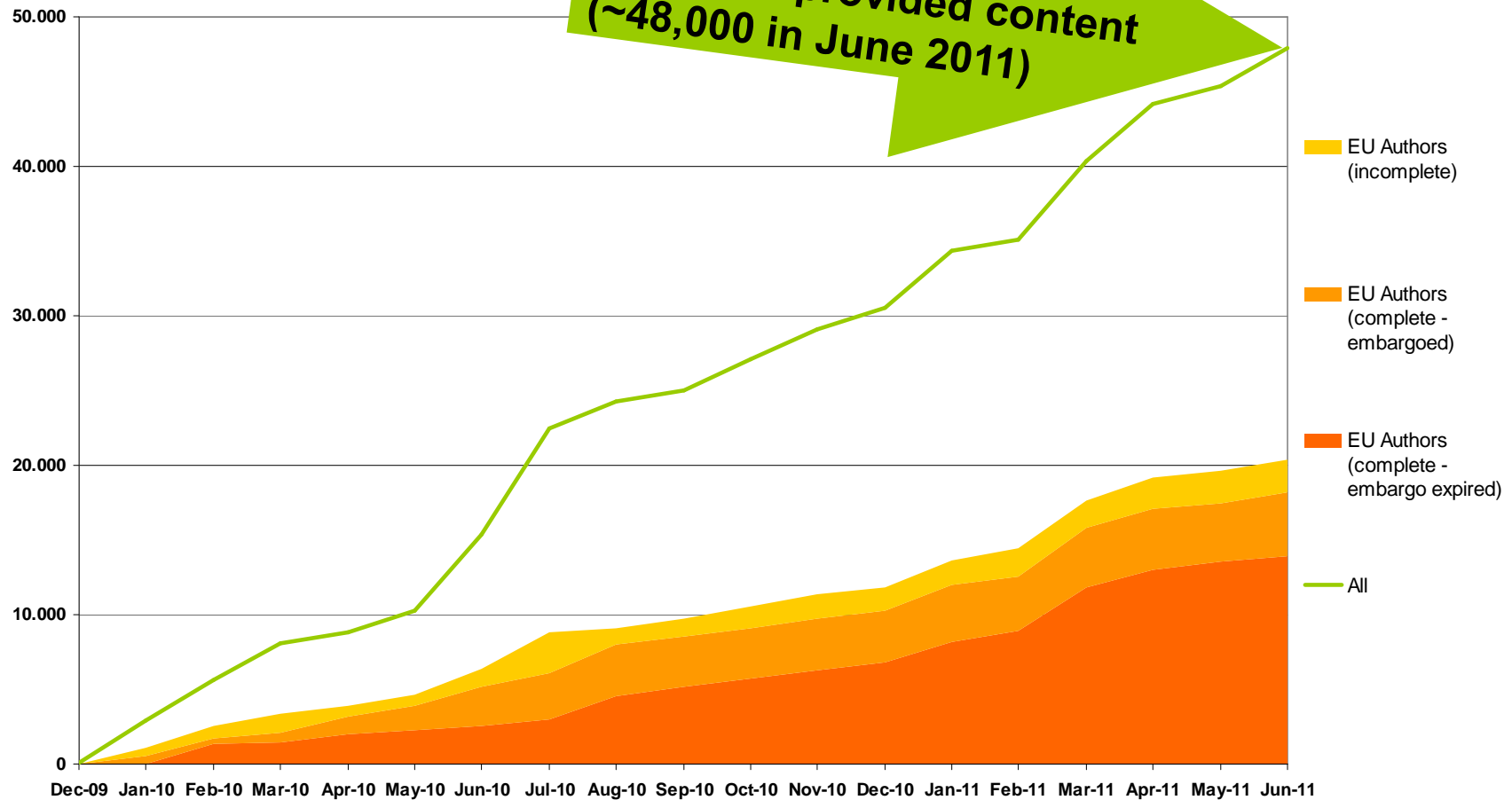
- The observatory consists of the
  - PEER depot
  - PEER repositories
- The depot
  - Acts as a „Clearing House“
  - Processes deposits
  - is a Dark archive
  - Distributes content to participating repositories
  - Hosted at INRIA
- The PEER repositories
  - provide the usage data need by our research partner CIBER
- Content inflow
  - 241 journals from four broad areas; selecting process
  - ~48.000 articles processed; ~14.000 EU deposits with embargo expired
  - 2 ways of articles deposit: publisher-assisted / author self archiving

# The PEER Observatory

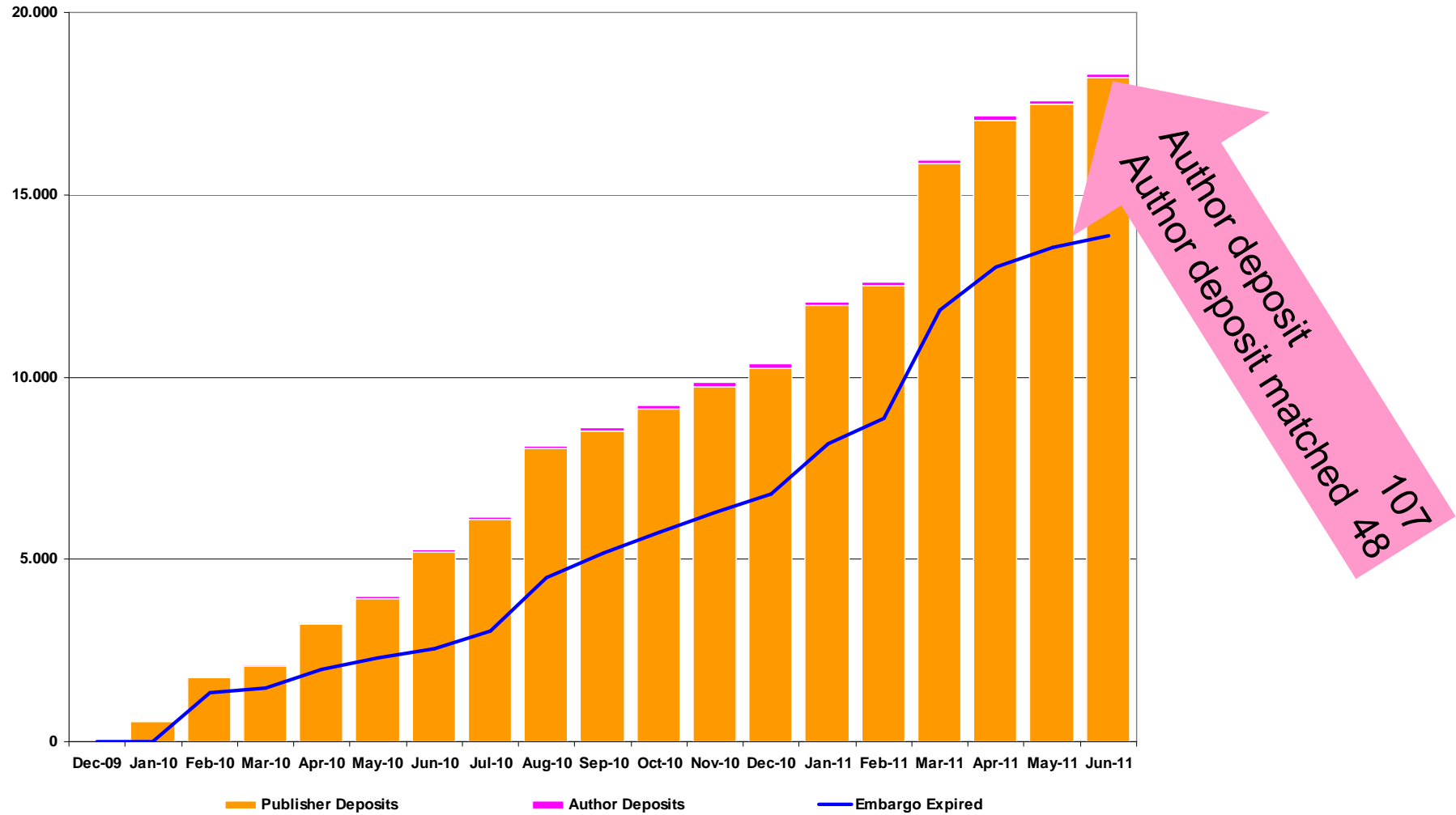
Develop an “observatory” to monitor the impact of systematically depositing stage-two outputs on a large scale



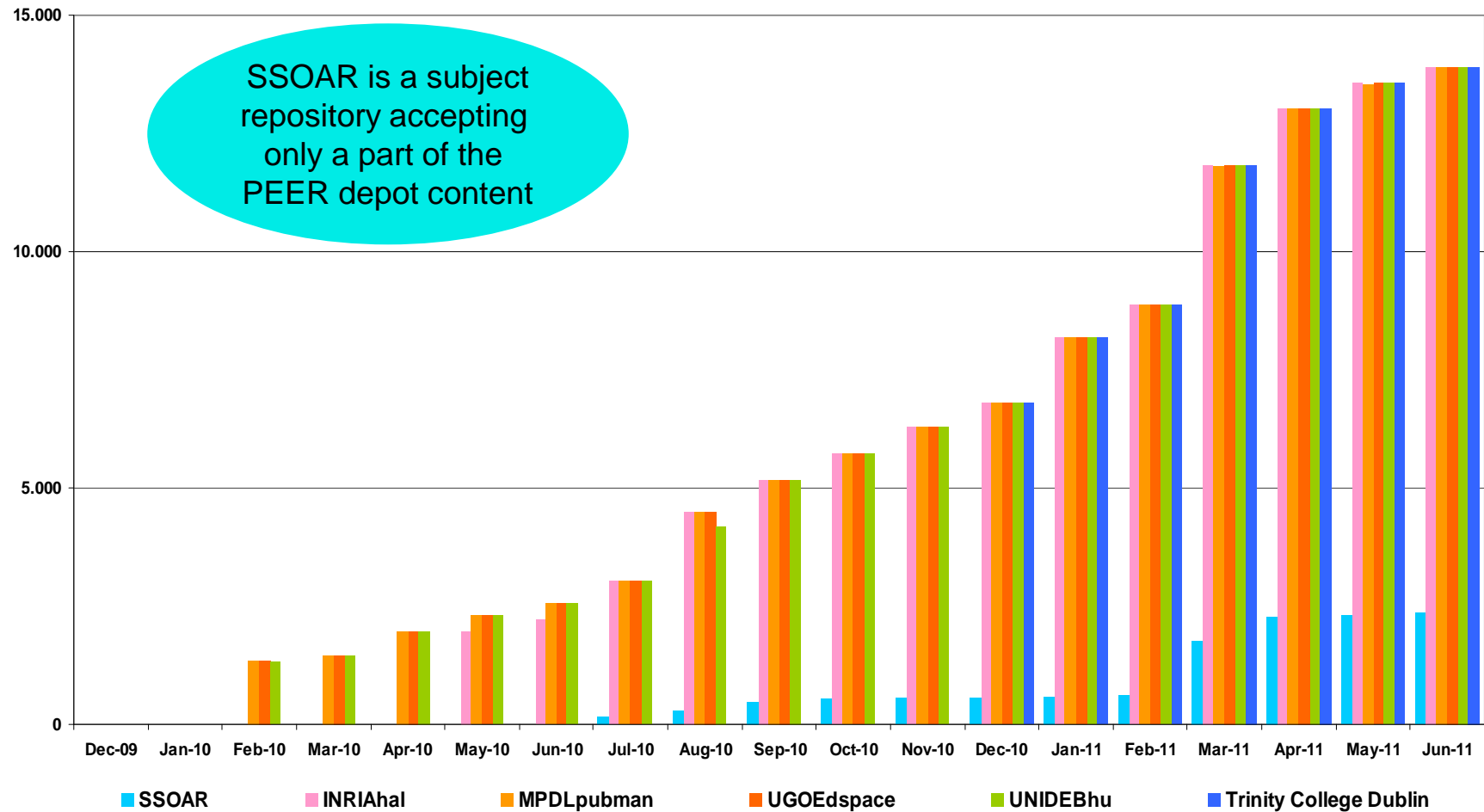
# Publisher deposits (cumulated)



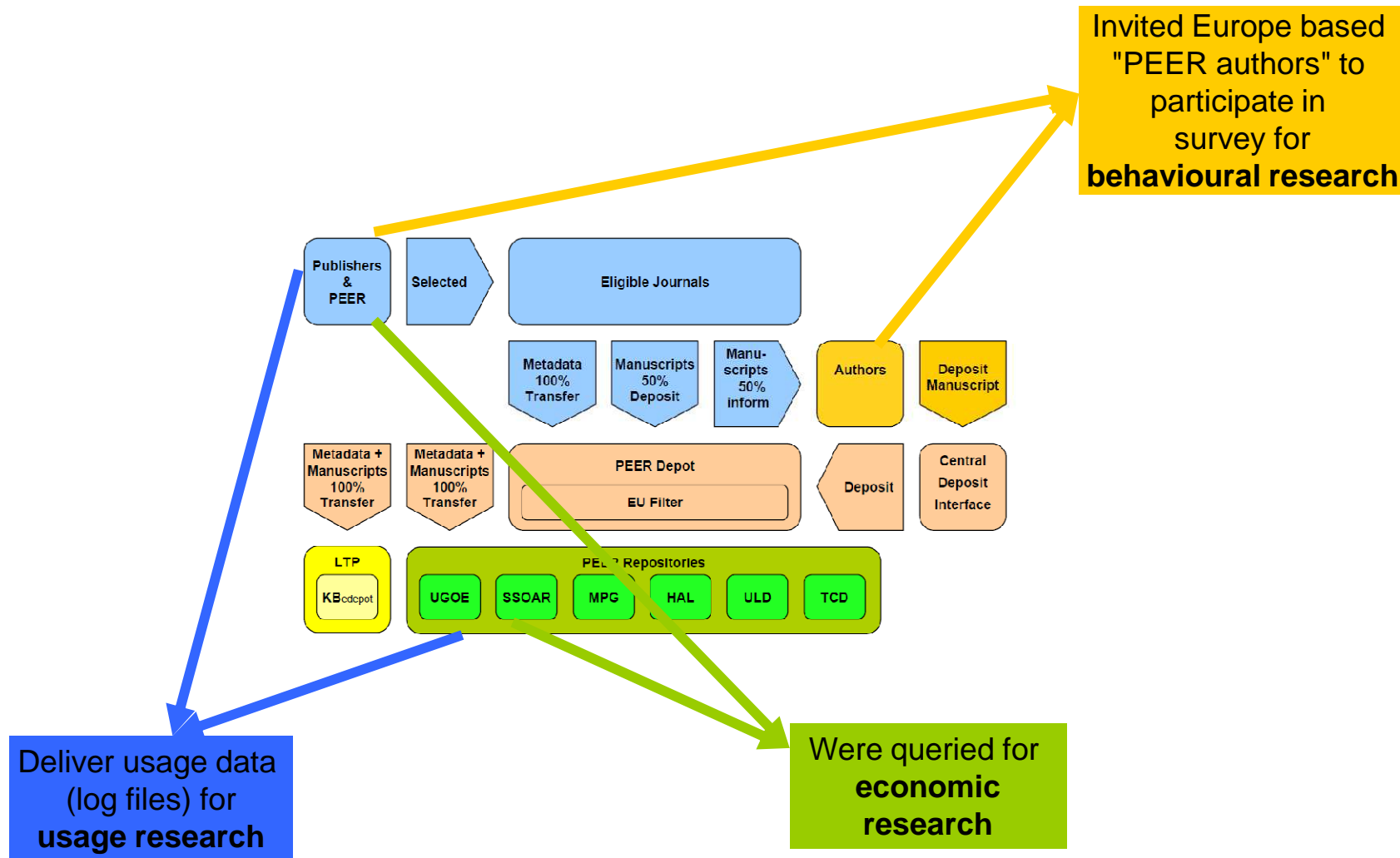
# EU-Deposits processed (cumulated)



# PEER repository population



# Peer Observatory + research projects



# Challenges and solutions (1)

## Challenges

- 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

## Solutions

- 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

# Challenges and solutions (2)

## Challenges

- 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

## Solutions

- Only one file format allowed – PDF
- Mapped into single TEI structure
- Extraction done at PEER Depot (GroBID)
- Changes, adjustments at publishers



# Challenges and Solutions – Publishers

- Arrangement of data transfer
  - Face to face meetings; first year of project (every 2 months), included technical representatives of PEER
  - Collective teleconferences
  - Publisher specific communications with PEER project manager
- STM - publisher cooperation
  - Quarterly strategic publisher meetings (face to face or telco)
    - report on progress
    - request for additional content
      - Additional journals
      - More back content (=embargo period expired)
      - Reduced embargo periods
- Technicalities
  - Some publishers amended their workflows to extract „accepted manuscripts“ & matching metadata; provided additional elements; transformed files to PDF

# Achievements to date

- 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  
Appointment and first results from all three research teams

# PEER research projects

- Usage research
  - CIBER group, University College London, UK
  - Objectives:
    - Determine usage trends at publishers and repositories;
    - Understand source and nature of use of deposited manuscripts in repositories;
- Economic research
  - ASK, Bocconi University, Italy
  - Objectives
    - Analyze the overall effects of large-scale Green OA on the economics of scholarly communication.
    - Investigate the cost of the large-scale deposit of stage-two research outputs; including the economic efficiency or cost of the process of deposit.
- 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.

# Behavioural research: Survey

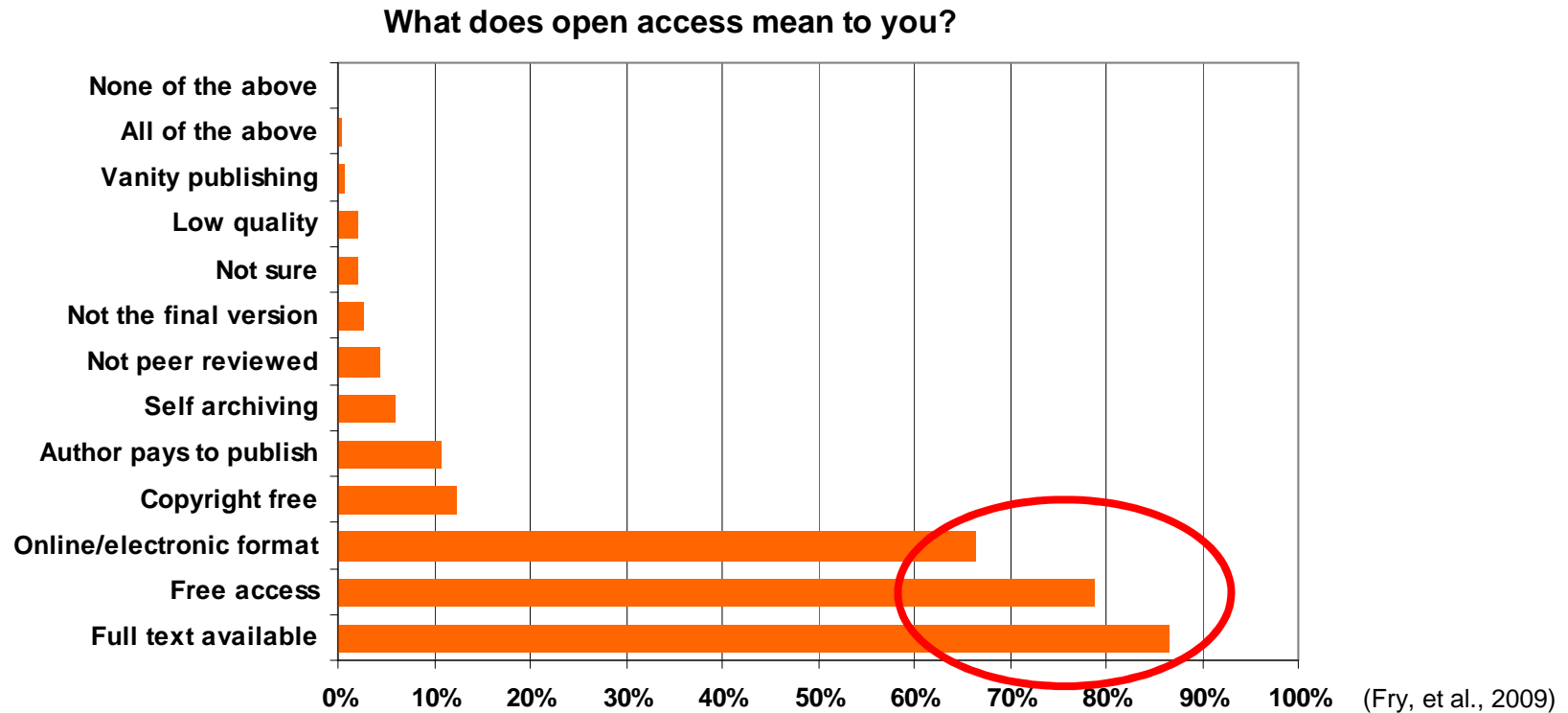
- Electronic survey of researchers conducted in Summer 2009.
- Invitations were distributed via the 12 publishers participating in the PEER Observatory.
- Distribution was restricted to EU-based corresponding authors, who had manuscripts published in those journals included in the PEER Observatory and the control group since the beginning of 2008.
- Non-EU respondents were filtered out from the responses in the analysis.



(Fry, et al., 2009)

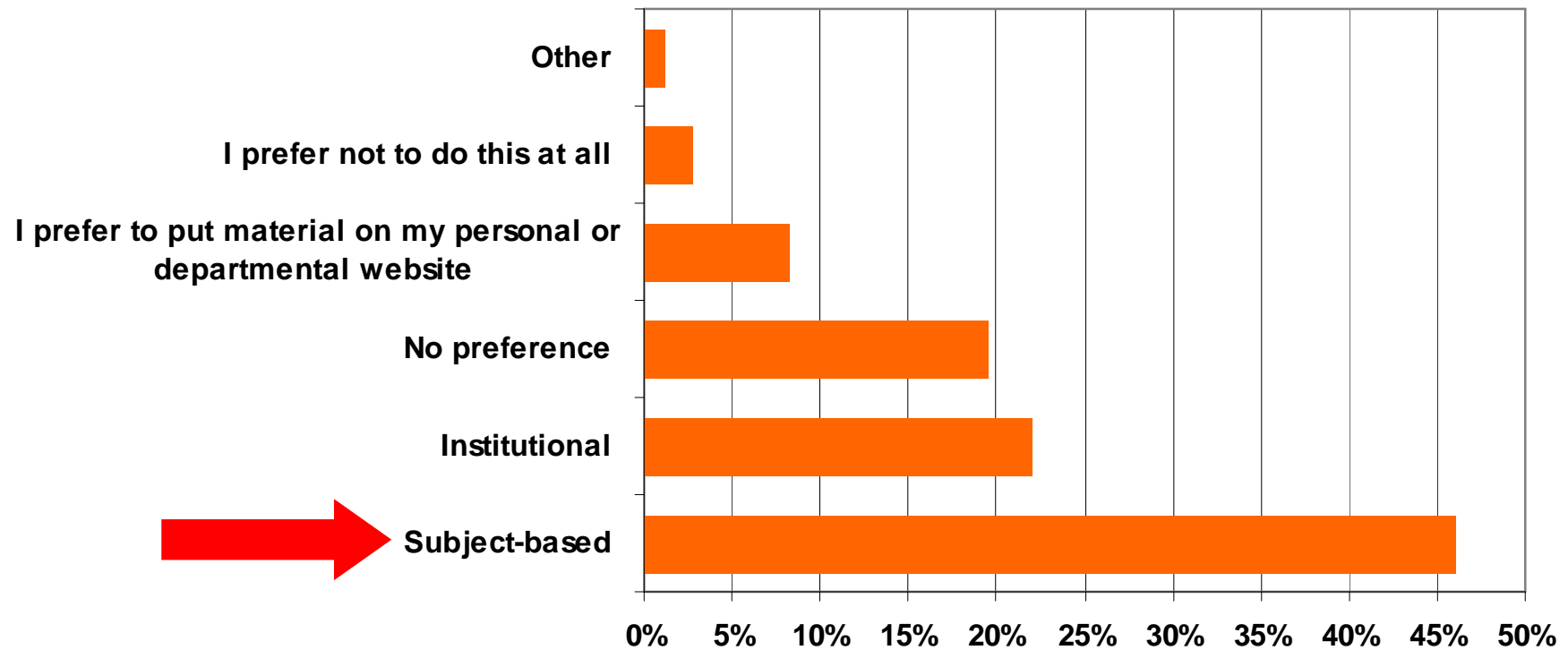
# Baseline survey: Understanding of Open Access

- More than two-thirds of respondents indicated that they understood open access to mean free electronic access to the full text of articles (30)



# Baseline survey: Preference for subject repositories

When considering placing material in publicly available repositories, do you have a preference for the type of repository?

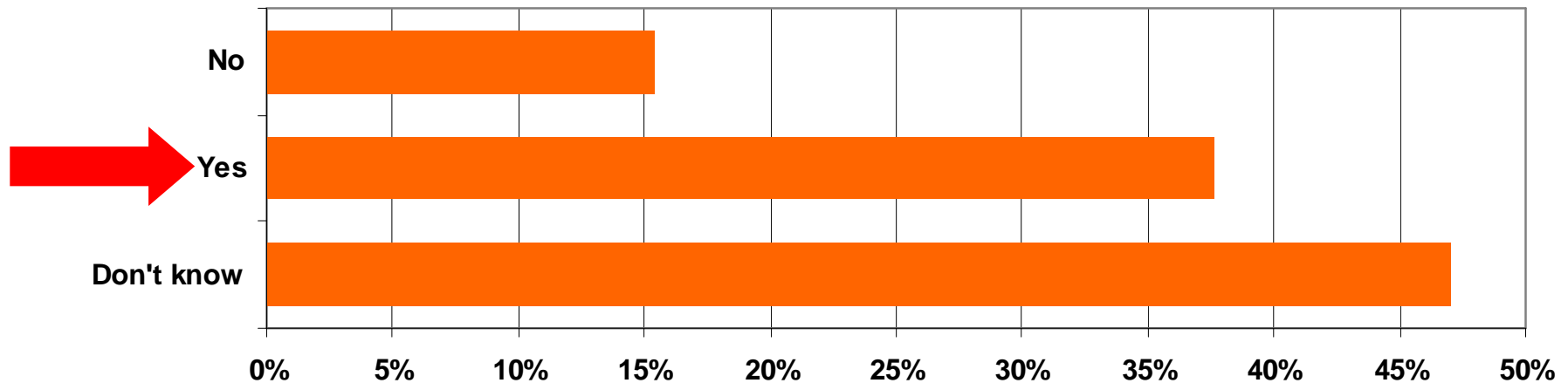


(Fry, et al., 2009)

# Baseline survey: Suitable subject repository available?

- Only 37% of responding authors think there is a subject repository suitable for their work available even though 46% favoured deposit in such a repository.

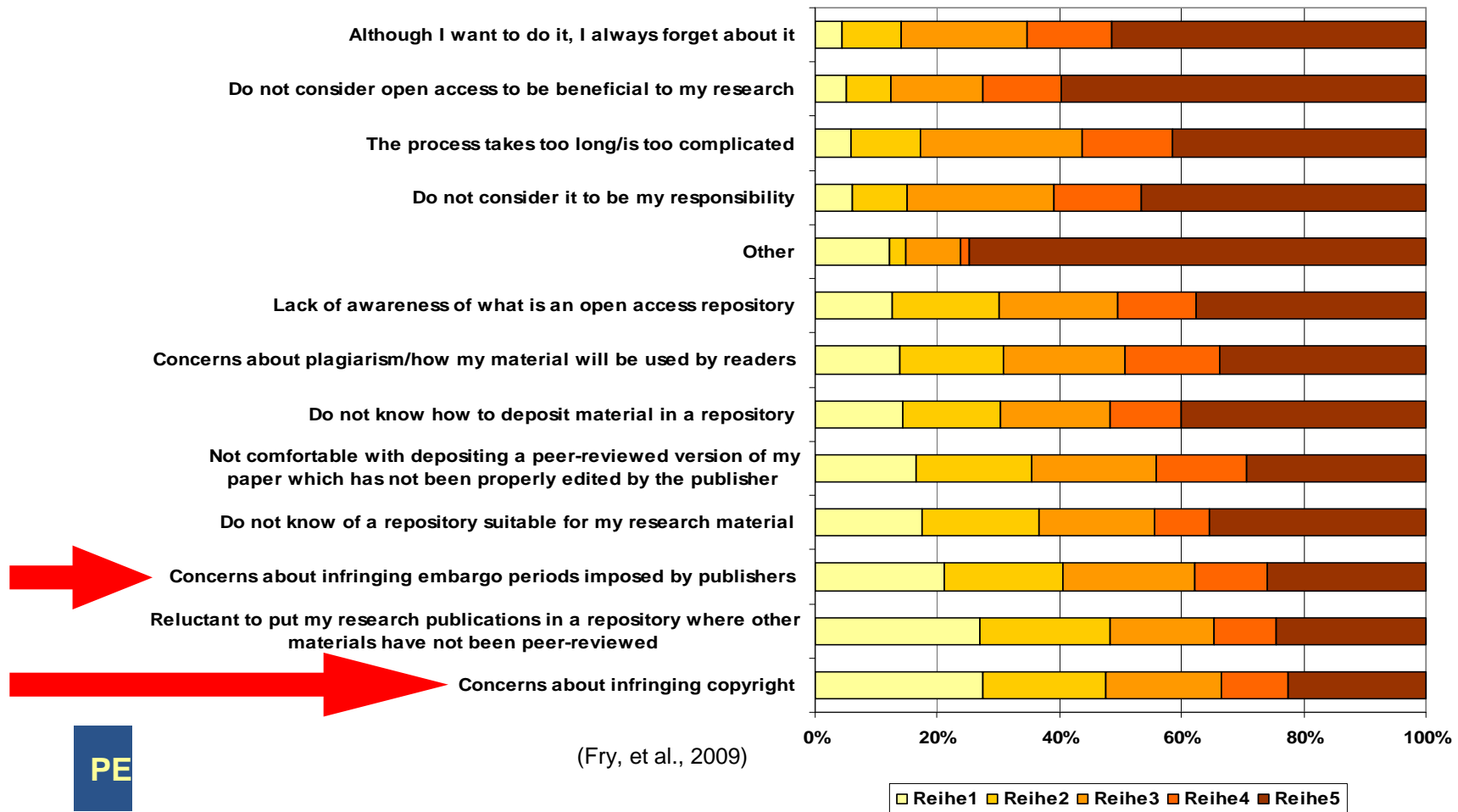
Are there any subject-based publicly available repository(ies) suitable for your research?



(Fry, et al., 2009)

# Baseline survey: Copyright concerns main problem

What reservations do you have about placing your peer-reviewed journal articles in publicly available repositories?  
(please rate on a scale 1-5, 1 = very important and 5 = not important at all)



(Fry, et al., 2009)



**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)**

# Sources

- Fry, Jenny et al. (IX. 2009): PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories. Baseline report.