



HAL
open science

Curating academic publications - a perspective for research libraries

Laurent Romary

► **To cite this version:**

Laurent Romary. Curating academic publications - a perspective for research libraries. ALPSP International Conference 2010, Sep 2010, Wyboston, United Kingdom. inria-00517052

HAL Id: inria-00517052

<https://hal.inria.fr/inria-00517052>

Submitted on 13 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Curating academic publications *a perspective for research libraries*

Laurent Romary
INRIA & HUB-IDSL

Why this talk?

- Background
 - Involvement in scientific information management in several major European research institution
 - CNRS, MPS, INRIA
 - Specific interest in content management and standards
 - TEI, ISO TC 37/SC 4
- General objective (bias)
 - Accompanying/anticipating the evolution of the publishing landscape – from the point of view of research institutions
- Issue
 - Is there a place for research libraries?...

Word of caution

- Research libraries come in many different forms
 - Heterogeneity across disciplines
 - The more libraries you see the less you may want to generalize
- Some research libraries have taken the lead in developing new technologies at the service of science
- Librarians come with a wide variety of profiles: information science, scientific background, computing

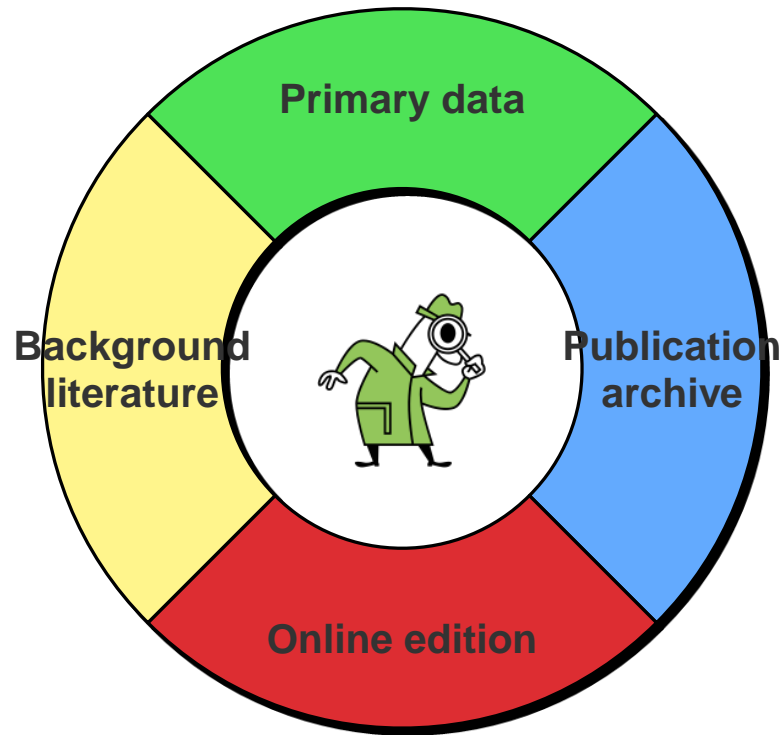
The good old library - missions

- Acquiring
 - Intermediary between academics and publishers
 - *But, not in the loop on the publishing side*
- Cataloguing
 - Managing, complementing meta-data
 - *But, not necessarily considered as an interlocutor by academics to retrieve information*
- Making available
 - Filing books and journals on shelves
 - *But, empty libraries (with a contrast between natural vs. human sciences)*

The library in the digital world

- Most assets have become digital
 - E.g.: subscriptions to scholarly publications, eBooks catalogues, databases etc.
- Business as usual
 - Acquiring
 - But still not in the loop of the publishing workflow
 - Cataloguing
 - But researchers use Google
 - Making available...
 - Online access is transparent to the reader
- Issues
 - Very little insight in the actual content
 - The academics have also become digital

The Scientist's (digital) ecology



Scientific information workflow

Working with research data

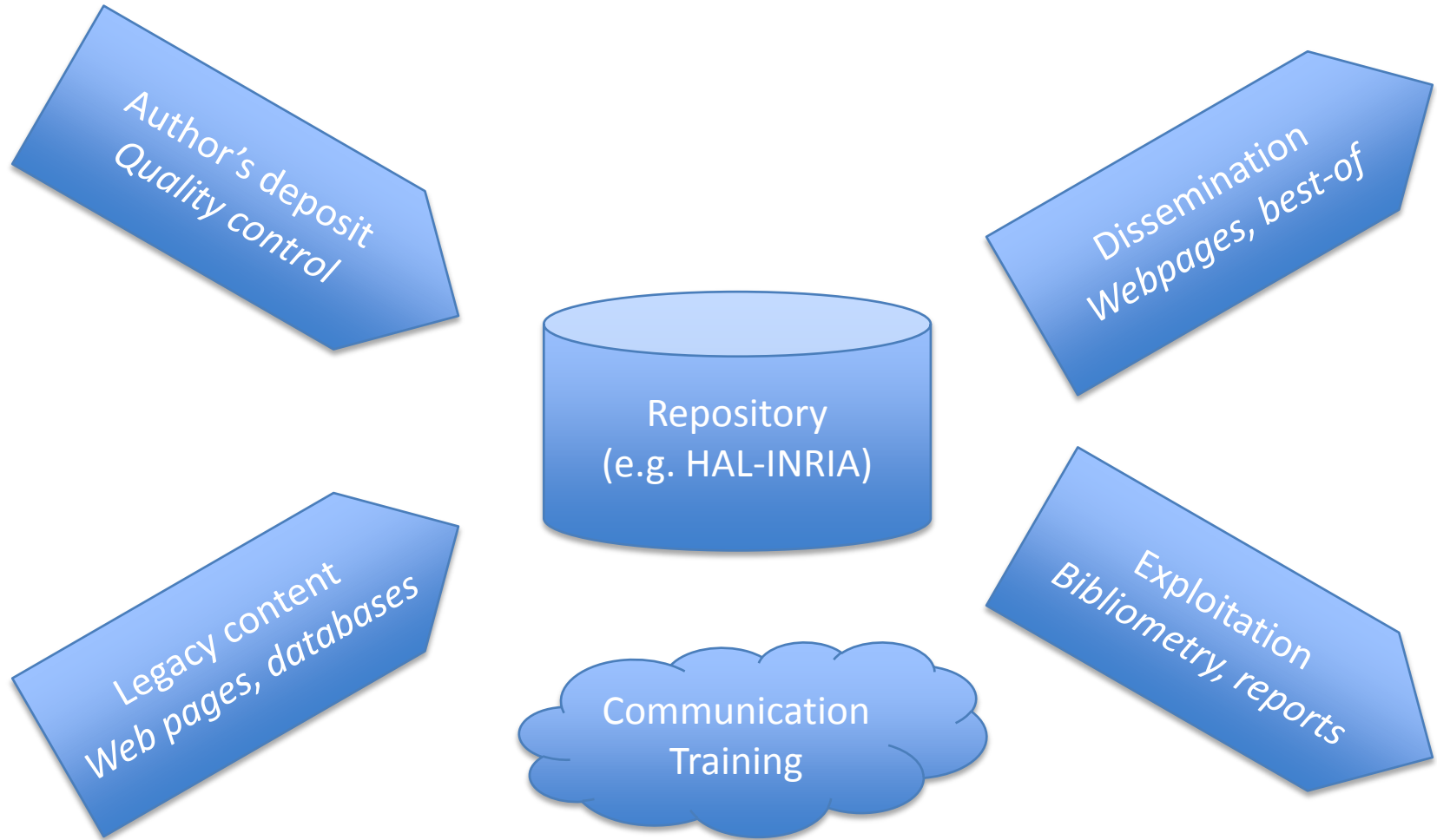
- Wide variety and complexity
 - High Energy Physics
 - Particle accelerators / colliders
 - Meteorology
 - Computer simulations
 - Astrophysics
 - Observations, stellar object descriptions
 - Biology
 - Spectrographic representations
 - Linguistics
 - Corpora, grammars, lexical databases
- What should be the role of research libraries?
 - Maintaining the link with publications?
 - Are research data too complex for them?

Consequence for the library

- Academics cannot master all this
 - Even if they think they can
- New needs are piling up
 - Open access
 - Putting together and managing publication repositories
 - Digitizing existing assets
 - Participating in the creation of content
 - Managing born digital documents
 - Digital publishing platforms, digital laboratory reports
 - Managing research data and primary sources?
- Not just metadata, management of content

Publication repositories

the editorial touch



The need for a more coherent infrastructure

- An old story – the MaxPlanck Society-Springer agreement (2008-2009)
 - Full open access scheme at publication time
 - Upload of Publisher's version onto the MPS repository
- Où les difficultés commencent...
 - Mapping the Springer and MPS technical platforms
 - Two evolving and at times incompatible digital environments

Why is it so difficult?

- Great heterogeneity of format within publishers
 - Meta data (and full-text)
 - Proprietary, ScholarOne, NLM 2.0, NLM 3.0, ...
- Various issues
 - Affiliations
 - Publication date information
 - ISO 639 codes (countries)
 - Bibliographical references
 - Proprietary metadata fields

The information chaos

- Article title
 - article-title/title | ArticleTitle | article-title | ce:title | art_title | article_title | nihms-submit/title | ArticleTitle/Title | ChapterTitle
- Journal title
 - j-title | JournalTitle | full_journal_title | jrn_title | journal-title
- ISSN (print)
 - JournalPrintISSN | issn[@issn_type='print'] | issn[@pub-type='ppub'] | PrintISSN | issn-paper
- First page of a paper
 - spn | FirstPage | ArticleFirstPage | fpage | first-page

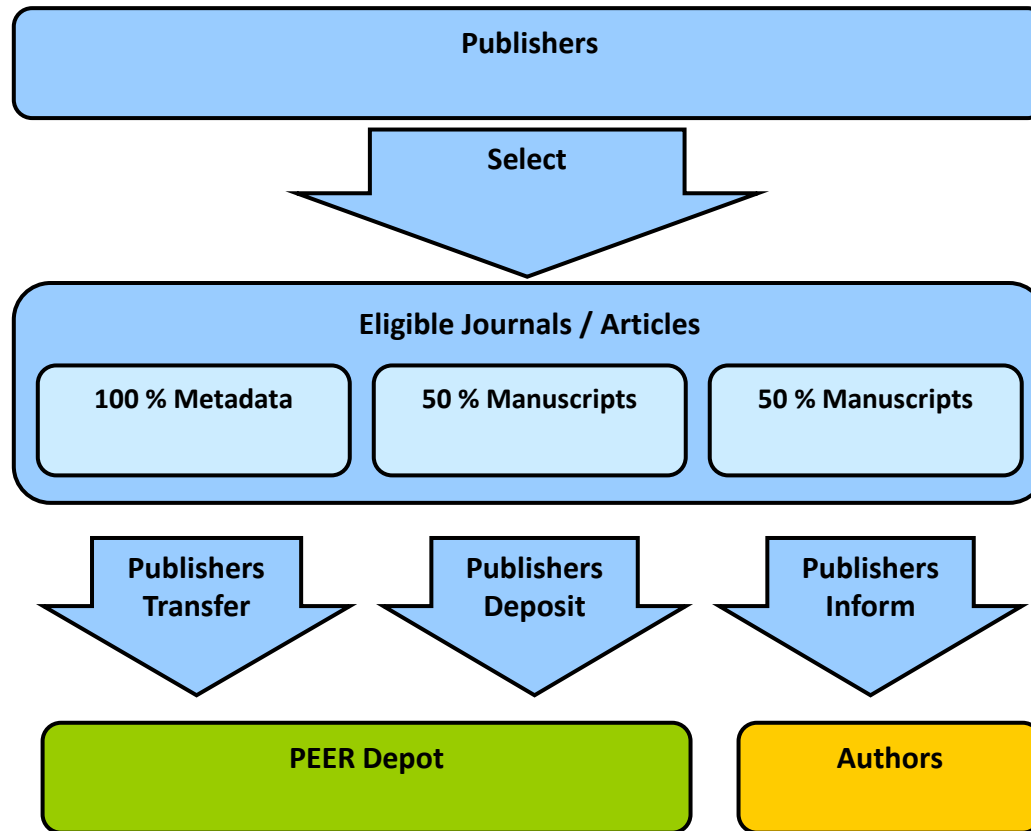
Sorting this out

- Defining a coherent infrastructure to facilitate
 - The long-term management of scholarly content in research institutions
 - Smooth interaction between publishers and research institutions
 - Better understanding of what each of us can provide
- On-going experimental setting: the EU PEER project

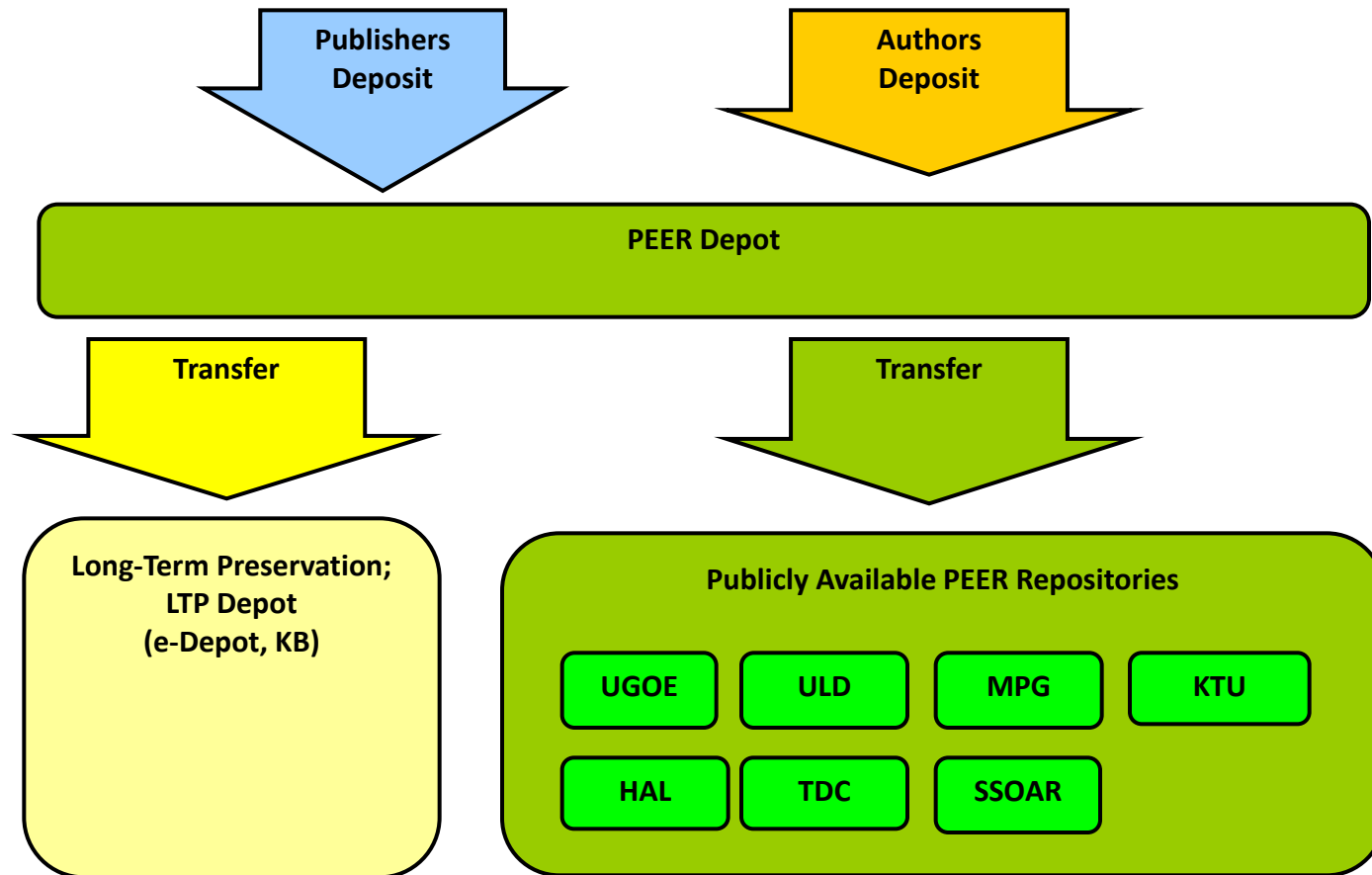
The PEER project

- Initiated by the EU commission (DG INFSO)
- Objective: study the impact of systematically archiving stage-two outputs in “institutional repositories” (cf. Romary & Armbruster 2010)
 - on journals and business models
 - on wider ecology of scientific research
- Consortium
 - STM, European Science Foundation (ESF), Goettingen State and University Library (UGOE), Max Planck Gesellschaft (MPG), INRIA

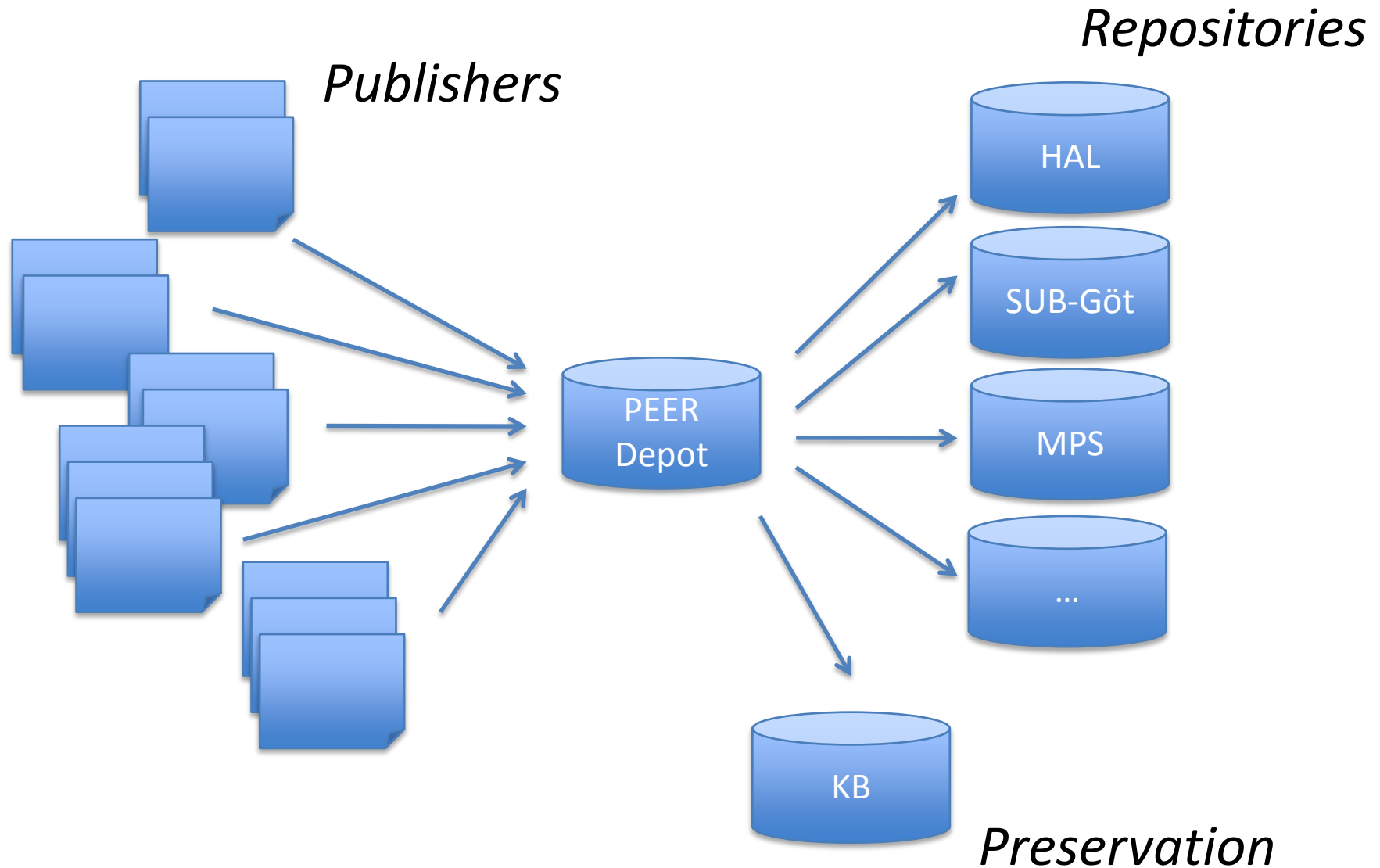
Content submission - publishers



Content submission – to repositories & LTP archive



The PEER deposit workflow



Consequence of the PEER experiments

- Strong involvement of librarians at INRIA
 - Defining workflows
 - Defining an optimal target format
 - Quality control (affiliations)
 - Perspective for full-text archiving
- Importance of a well-defined standardization strategy
 - TEI – Text Encoding Initiative
 - Well documented and stable initiative
 - Clear maintenance strategy
 - Systemic view of digital documents
 - Articles, books, primary sources
 - Large community of practice
 - Scholarly publishing
- A continuity with existing research library skills

Conclusion

- New profiles
 - Digital curators
 - Combining old skills with new technologies
 - Describing digital sources: meta-data
 - Understanding and representing the structure of digital sources
 - Enriching (annotations, links), versioning, disseminating
 - Proximity to researchers
 - Higher profiles – understanding scholarly work
 - Flexibility
 - Need to adapt to scenarios which are not yet anticipated
- New roles – articulating research and publishing
 - Cf. PEER
 - Management of (“gold”) OA deals
 - Long term preservation
 - Support for scientific data management
 - Contribution to the evolution of scholarly communication

Further readings

- BIJSTERBOSCH, M., BRÉTEL, F., BULATOVIC, N., PETERS, D., VANDER-FEESTEN, M., WALLACE, J., PEER. D3.1. Guidelines for Publishers and Repository Managers on Deposit, Assisted Deposit and Self-Archiving, 2009.
http://www.peerproject.eu/fileadmin/media/reports/D3_1_Guidelines_v8.3_20090528.Final.pdf
- HOLMES, M., ROMARY, L., “Encoding models for scholarly literature”, in KAPIDAKIS, S. (dir.), *Publishing and Digital Libraries: Legal and Organizational Issues*, 2009.
<http://hal.archives-ouvertes.fr/hal-00390966/fr>
- ROMARY, L., ARMBRUSTER, C., “Beyond institutional repositories”, *International Journal of Digital Library Systems*, 2010.
<http://hal.archives-ouvertes.fr/hal-00399881/fr>