

# Repositories – Practice & Vision

## Overview 2010

# This Talk

- Objective
  - Repositories on the Menu 2010
  - Overview of a highly diversified field as compared to 2005
- Coverage
  - selective, subjective, observational, cursory, data-centric
- Audience addressed
  - Repository Managers, Information Managers, Funders ...
- Approach wisdom of crowds
  - Open Repositories 2010 Conf. assumed as representative
  - General session themes taken as starting points

# Open Access Policy

- **Vision: Open Access objects as commodity**
  - Not an ethical issue but most efficient mode of scholarly and scientific communication
- **Practice: Considerable growth but tedious**
  - More policies (publisher; institutional; funders: EC)
  - More repositories & documents (cf. BASE/Sherpa)
  - But systematic input of data lacks automation
- **Forecast: Further incremental progress**

# Bibliography & Citation

- Vision: R's for authors' publication management
  - Personal publication lists & reference management
- Practice: Please full-texts – but contextualized
  - Completion is handwork, augmented by Web2tools
  - Citations not in R's: GS, WoS, Scopus, PMC, Inspire
- Forecast: Completeness is clue
  - Authors only take serious if really good

# Usage Statistics

- Vision: Scholarly metric alternative to citations
  - Higher precision and resolution
  - Also important for service optimization „Research Intelligence“
- Practice: Still measured non-standardized
  - Comparison and integration impossible
  - Pilots underway: (Cross-)National, EU, Publisher
- Forecast: Needs a few years but will fly

# Administrative Systems

- Vision: Authors use repositories for reporting
  - Institutional bibliography, policy compliance
  - Why report twice? Double-input a „no-go“!
- Practice: Cautious opening of „official“ systems
  - Re-use – e.g. alerting – requires disambiguation
  - Reporting needs authoritative data persons, departments...
  - But: closed shops and IT-integration problems e.g. IDM
- Forecast: Success for R's with CRIS interfaces

# Data Integration & Disambiguation

- **Vision: Automated exchange of publication data**
  - Bibliographic data is already out there: so re-use!
- **Practice: Examples? Yes, but few!**
  - Fetch of data from WoS, ArXiv, PMC: Who does it?
  - E.g. Author disambiguation: „Is Meyer Meyer?“
  - Also departments, institutions, classifications
- **Forecast: (Inter-)National authority files /IDs**
  - Terminology services used to disambiguate

# National Approaches

- **Vision: Institutional>national>global structure**
  - Geographic regime builds on reliable institutions
- **Practice: Europe strong, but all is still stringy**
  - Some functional chains of comm. OpenAIRE, Japan, China
  - Disciplinary worlds (and US) work differently
- **Forecast: Mixed geographic and subject system**
  - Invisible automatic exchange will blur boundaries

# Research Data

- Vision: Researchers *can* easily submit data
  - Higher transparency, cost-efficiency
  - But please provide differential access
- Practice: Highly heterogeneous over disciplines
  - E.g. Bioinformatics (~open) & engineering (~closed)
  - Funders require data management plan
- Forecast: Subject based practice dominates
  - No generic basis as in text-based documents
  - Only basic archiving may be common denominator

# Digital Preservation & Archiving

- Vision: Researchers have trusted archives
  - Particularly important for research data, where no est. publishing system can catch the trust
- Practice: Documents feasible, Data in question
  - Many important insights, but too ambitious
  - National libraries and data archives on the move
- Forecast: hot-service if approached practically
  - 3-5-10 years pragmatic preservation: then let's see!
  - „What can you promise?“ vs. „What can't be promised“

# Repository Platforms

- Vision: Solid and sustained multi-purpose R's
  - Research Data and administrative functionality
  - Developer communities go business
- Practice: Extension of existing, still newcomers
  - Dspace, Eprints, Fedora, Opus extended
  - PubMan/eSciDoc, Invenio etc. maturing
  - „Competition“ from VCSs and Web-CMS?
- Forecast: Further consolidation

# Repository Interoperability

- Vision: R's with many tentacles and glands
  - Bidirectional interfaces for data and protocols
  - „Repositories as pipelines“
- Practice: OAI-PMH still dominates, others come
  - REST/SOAP, XMLHttpRequest, SWORD, SRU, Feeds, OAI-ORE, Linked Data / RDF: what have you got!?
- Forecast: Technology savvy staff required
  - Customizing interfaces not out-of-the-box

# Academic Workflows

- Vision: R's embedded in research process
  - Virtual Research Environments: Service mash-ups
  - Specificity requires customization
- Practice: Again, docs feasible – data in question
  - Generic: Office, calendar, web-sites, documentation
  - Specific: Research tools and workflows
- Forecast: Self-maintained repository platforms
  - Research cultures vs. institutional service dev.
  - Local tekkies and central service collaboration

# Sustainability & Business Operations

- **Vision: Balanced public and private businesses**
  - Institutional services as backbone that sources out
- **Practice: Libraries dominate – Cloud coming up**
  - Separation between R's and publishers blurs slightly
  - Research data archiving presses computing centres for novel virtualization strategies
- **Forecast: Mixed institutional & outsourced op's**
  - Good reasons for keeping data at home will level with price considerations for comm. cloud services

# Repository Infrastructures

- Vision: R's data from the socket outlet
  - Researchers and tekkies just have to plug-in
- Practice: R's & infrastructural systems converge
  - Cautious attempts to go beyond OAI-PMH
  - Links to CRISs and disciplinary systems
  - Infrastructures allow data and service mash-ups
- Forecast: The content network tightens
  - R's without user interfaces; R's as data providers

# Take Home Message(s)

- R's 2010
  - Are not only about Open Access anymore
    - Successive publication stages of texts & research data
  - Extend towards administration and research data
    - Bibliographic re-use, Disambiguation/IDs
  - Are embedded and embed themselves
    - Academic workflows / VREs, Frameworks, Infrastructures
  - Become an asset vessel of the institution