Technologies & Tools for the Implementation of Repositories and E-journals

Dr Nikos Houssos
Head of Software Development Unit, EKT / NHRF
www.ekt.gr

International Conference: "Open Access Infrastructures: The Future of Scientific Communication"

15-16 December 2008, Athens-Greece (National Hellenic Research Foundation)
Agenda

- Implementation options and challenges repositories and e-journals
- What is already available
- What needs to be improved
Repositories and E-Journals Implementation options

• Host in-house
  – Develop on open-source platform
  – Develop on commercial platform
  – Develop from scratch

• Outsource hosting

• Platform selection
Open source platforms

• Flexibility, unlimited customisation capabilities
• Tested solution - support already thousands of open access systems installations worldwide
• Suitable for a wide range of implementations
  – From the smallest to the largest
  – Not only for high-end organisations at the edge of technology
Systems infrastructure requirements

- Reliability / availability / safety
- Rapid response to changing requirements
  - On-demand, “late” resource allocations
  - Staging environments - virtualisation
  - Particularly important for development on open source
- Logging / reporting
Systems infrastructure issues

- Virtualisation management
- Storage management
- Monitoring availability and performance
- Logging and statistics
The Greek website for open access is part of the project “National Information System for Research and Technology, Phase III – Open Access Electronic Repositories and Journals” which is being implemented by the National Documentation Centre under the framework of “Digital Greece” (www.psifakiellada.gr) and is co-funded by the European Union - European Regional Development Fund (80%) and by the Hellenic State (20%) through the Operational Programme Information Society (3rd CSF 2000-2006).
Software infrastructure issues

• Robust open source platforms are already in place for repositories and e-journals
• Basic functionality is well supported
• Some important advanced features are still not widely available
Mature functions

- Indexing of metadata and full-text
- Metadata search and browse
- Support for fundamental workflows (e.g., content submission, peer review)
- Protocols for remote metadata harvesting and searching
- Basic multi-lingual material support
Topics for further consideration (1 / 2)

• Advanced mechanisms for content submission
• Enhancing the end-user experience - web 2.0
• Sophisticated metadata representation standards (e.g., for compound objects)
• Semantic search / retrieval
• Applications of data matching
Topics for further consideration (2 / 2)

- Usability of full-text search (e.g., highlighting)
- Protocols and APIs for remote updates and harvesting full-text
- Distributed workflows
- Single-sign on
- Interoperability of repositories, e-journals and CRIS
How to attract researchers?

• Improve the user experience of researchers in the role of data providers
  – Help them with loading data to repositories

• Provide value-added services over repositories and e-journals
  – Reference management
  – Reporting
  – Usage statistics of researcher output
Researchers as data providers

- Combination of multiple mechanisms to make things easier for the researcher
  - Loading and pre-processing of researcher data from multiple bibliographic sources
  - Integration with workflows for production of scientific documents - “zero-click ingest”
  - Automatic metadata extraction
- Several production systems in use worldwide
- No open source solution yet
The end user experience

- Web 2.0 – like features
- Faceted browsing
- Maps, time lines, time plots
- Social networking features
  - Tagging, reviews
- Personalisation
- Make the above cross-repository?
The Greek website for open access is part of the project "National Information System for Research and Technology, Phase III – Open Access Electronic Repositories and Journals" which is being implemented by the National Documentation Centre under the framework of "Digital Greece" (www.psifiakiellada.gr) and is co-funded by the European Union - European Regional Development Fund (80%) and by the Hellenic State (20%) through the Operational Programme Information Society (3rd CSF 2000-2006).
Metadata representation

• Current systems support mostly flat metadata standards
• Limited support for:
  – Heterogeneity in metadata standards – model-driven development
  – Compound objects
  – Linking
The significance of data matching

• Authority files / unique identifiers for names (authors, research organisations, journals, publishers, research areas)

• Benefits:
  – Enhanced search / browse functionality
  – Reliability in statistics and reporting
  – Easier data integration and consolidation
The Greek website for open access is part of the project “National Information System for Research and Technology, Phase III – Open Access Electronic Repositories and Journals” which is being implemented by the National Documentation Centre under the framework of “Digital Greece” (www.psifiakiellada.gr) and is co-funded by the European Union - European Regional Development Fund (80%) and by the Hellenic State (20%) through the Operational Programme Information Society (3rd CSF 2000-2006).
Example: data matching

data matching list washing object identity resolution entity disambiguation coreference resolution reference reconciliation duplicate detection merge/purge processing data deduplication instance identification record matching database hardening name matching identity uncertainty entity resolution entity matching record linkage

The Greek website for open access is part of the project “National Information System for Research and Technology, Phase III – Open Access Electronic Repositories and Journals” which is being implemented by the National Documentation Centre under the framework of “Digital Greece” (www.psifiakiellada.gr) and is co-funded by the European Union - European Regional Development Fund (80%) and by the Hellenic State (20%) through the Operational Programme Information Society (3rd CSF 2000-2006).
Interoperability standards

- Interoperability guidelines for repositories (DRIVER II, DINI, ...)
- Harvesting standard (OAI-PMH)
- Search standards (Z39.50, SRW/U)
- Structure (METS, OAI-ORE, MPEG-DIDL)
- Batch remote submission (SWORD)
- Need for standard update APIs
Conclusions

• Open access installations have demanding requirements on the systems infrastructure
• Robust open source software platforms are available for both repositories and e-journals
• Fundamental functionality is well supported
• Further work needed for several important advanced functions / features
The Greek website for open access is part of the project “National Information System for Research and Technology, Phase III – Open Access Electronic Repositories and Journals” which is being implemented by the National Documentation Centre under the framework of “Digital Greece” (www.psifiakiellada.gr) and is co-funded by the European Union - European Regional Development Fund (80%) and by the Hellenic State (20%) through the Operational Programme Information Society (3rd CSF 2000-2006).

Back up slides
Technical enablers

• Systems
  – Virtualisation
  – Single sign-on

• Software
  – Web services / Service-oriented architecture
  – Model-driven development
  – User interface technologies
Multi-lingual information management

- Multi-lingual information management
  - Encoding
  - User interface localisation
  - Indexing, searching and browsing