



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)



ΕΘΝΙΚΟ ΚΕΝΤΡΟ
ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL
DOCUMENTATION
CENTRE



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.psiikiellada.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).



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



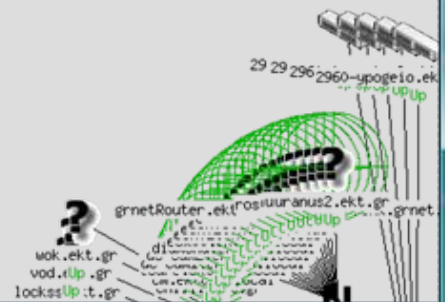
Systems infrastructure requirements

- 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



172.16.200.36 - VMware Infrastructure Client

File Edit View Inventory Administration Plugins Help

Hosts & Clusters

Overview

Statistics for openaccess.ekt.gr (2008-11) - main - Windows Internet Explorer

http://stats.ekt.gr/awstats/awstats.pl?month=11&year=2008&output=main&config=op

Statistics for openaccess.ekt.gr (2008-11) - main

Στατιστικά του: openaccess.ekt.gr

Περίληψη

Περίοδος αναφοράς: Μήνας Νοέ 2008

Πρώτη επίσκεψη: ΝΑ

Τελευταία επίσκεψη: 30 Νοέ 2008 - 23:58

Μοναδιαίοι επισκέπτες	Αριθμός επισκέψεων	Σελίδες	Επιτυχίες
Κίνηση που προβλήθηκε	10520	17571	79820
Κίνηση που δεν προβλήθηκε *			63362
		(4.54 Σελίδες/Επίσκεψη)	(22.54 Επιτυχίες/Επίσκεψη)
			396170

Edit View History Bookmarks Tools Help

http://cacti.ekt.gr/cacti/graph_view.php?action=preview&host_id=45&graph_template_id=0&filter=+

Host: Pandektis Template: Any Search: go clear

Refreshes: Last Day From: 2008-12-11 14:28 To: 2008-12-12 14:28 1 Day refresh clear

Showing Rows 1 to 10 of 12

Pandektis - CPU Usage

Pandektis - Detailed Memory

Pandektis - JVM GC Stats

Pandektis - JVM Heap Utilization



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.psi-fiakiellada.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

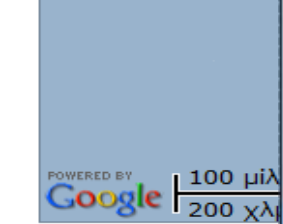
- Combination of multiple mechanisms to make things easier for the researcher
 - Loading and pre-processing of researcher data from multiple bibliographic sources
 - Integration with workflows for production of scientific documents - “zero-click ingest”
 - Automatic metadata extraction
- Several production 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?

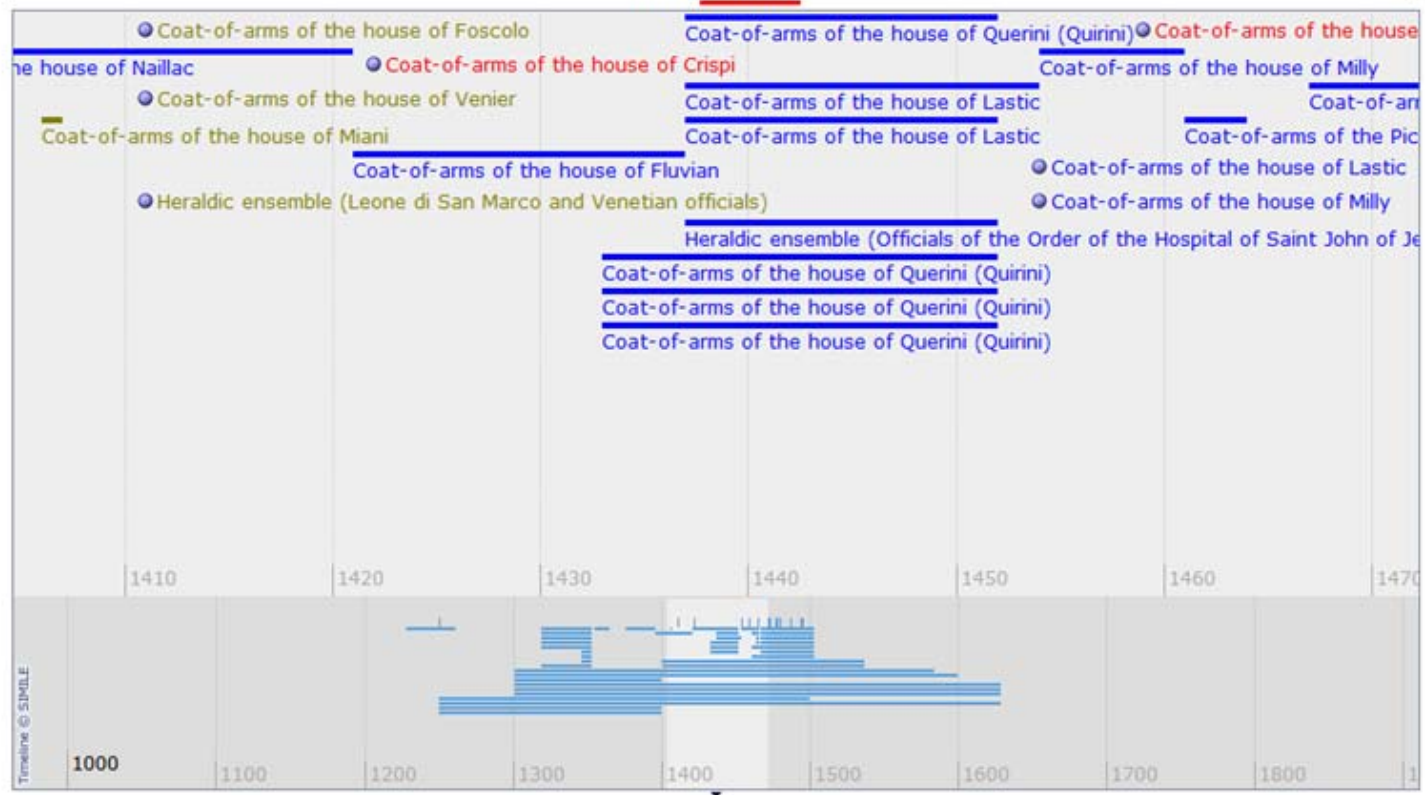
Heraldic data overview



- Region**
- 27 **Attiki**
 - 20 **Ionian islands**
 - 42 **Kyklades islands**
 - 2 **Peloponnes**

Heraldic data overview (faceted browsing)

MAP • TIMELINE



- Crete
- Dodecanese islands
- Kyklades islands
- Peloponnese

Region	4 <input checked="" type="checkbox"/>	Settlement	Historical Period	1 <input checked="" type="checkbox"/>
22 Crete	<input checked="" type="checkbox"/>	2 Geraki (Peloponnisos)	105 13th-15th century	<input checked="" type="checkbox"/>
69 Dodecanese islands	<input checked="" type="checkbox"/>	4 Kalymnos	111 16th-17th century	<input type="checkbox"/>
6 Kyklades islands	<input checked="" type="checkbox"/>	4 Kamariotis (Malevizi-Iraklio, island of Crete)	42 18th-19th century	<input type="checkbox"/>
9 Northern Aegean islands	<input type="checkbox"/>	4 Kastri of Agios Mamantas (Mylopotamos-Rethymno, island of Crete)	2 20th century	<input type="checkbox"/>
8 Peloponnese	<input checked="" type="checkbox"/>	4 Kefali (Kissamos-Hania, island of Crete)		
		1 Kefalos (island of Paros)		
		29 Kos (city)		



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



MADS Collection Editor

Αρχείο Γλώσσα Βοήθεια

Νέο Αρχείο
 Άνοιγμα
 Αποθήκευση
 Εξοδος

Λίστα Συγγραφέων

- Agriantoni, Christina
- Alexis, Michael N.
- Anagnostakis, Ilias
- Andrianou, Dimitra
- Andriotis, A. N.
- Angelidi, Christine
- Apostolopoulos, Dimitris
- Argyropoulos, Roxane
- Asdrachas, Spiros
- Aslanis, Ioannis
- Aspromallis, G.
- Athanasίου, K.
- Bacalis, Naoum C.
- Balta, Evangelia
- Bonas, G.
- Calogeropoulou, Theodora
- Cazianis, C.T.
- Cefalas, Alciviadis Constantinos
- Chatziioannou, Aristotelis
- Chatziioannou, Maria, Christine

Αλλαγή Στοιχείων

Όνόματα Προβολής

Γλώσσα	Όνομα
en	Chryssikos, Georgios D.
gr	Χρυσικός, Γεώργιος Δ.

Εναλλακτικά Ονόματα

Γλώσσα	Όνομα
en	Chryssikos, G.D.
en	Chryssikos, D. G.
en	Chryssikos, G.
en	Chryssikos, G. D.
en	Chryssikos, G.D.
gr	Χρυσικός, Γ.
gr	Χρυσικός, Γ. Δ.
gr	Χρυσικός, Γ.Δ.

Υπό Σκέψη

- Adamopoulos, G.
- Aggeli, A.
- Agiostratidou, G.
- Athanasίου, C.
- Benten, W. P. M.
- Berberian, J.
- Beretta, Marco
- Bergman, T.
- Bernardi, P.
- Bernet, B.
- Besnardramage, M. J.
- Besra, G. S.
- Bethell, D.
- Bex, V.
- Bezrukov, V.
- Bhat, A. G.
- Bia, M.
- Bianchini, F.
- Bibby, M. C.
- Biberacher, W.
- Richard, C. J. F.



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**



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

**Open
Access**

knowledge for all



Back up slides



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.psi-fiakiellada.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).



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