Pandektis: Implementing a repository of greek historical and cultural material with DSpace

Nikos Houssos

Ilias Stavrakis

Kostas Stamatis

Ioanna-Ourania Stathopoulou

Christina Paschou

National Documentation Centre / National Hellenic Research Foundation

Open Repositories 2008
Southampton, UK

03 April 2008

Agenda

- About the National Documentation Centre and open access activities
- About the Pandektis project
- Technical aspects of Pandektis implementation
- Further work

NDC activities and role

- Part of National Hellenic Research Foundation (NHRF) - oldest research centre in Greece - signatory of the Berlin Declaration on Open Access
- Research and technology Content:
  - National dissertation thesis archive
    - 13,000+ thesis on line, 3.500.000 pages in total
    - Both born digital and digitized material
  - National archive for variety of DBs and all research journals – both in print and electronic ones.
  - Union catalogues of journals that Greek Academic Libraries subscribe, municipal libraries etc.
  - Digitization services for cultural and scientific content
- Services for the academic/research community:
  - Research results dissemination, international cooperation, Enterprise Europe Network, National Contact Point for FP7, research metrics extraction

National Documentation Centre – open access related activities

- Since early 2007 runs funded projects related to open access infrastructures
  - Pandektis cultural collections
  - Develop the institutional repository of NHRF and make the first step towards an infrastructure for nationwide thematic repositories
  - Operate 6 open access journals in humanities and physical sciences
  - Promote open access movement awareness (openaccess.gr)
- Funding for developing open access systems is (practically) secured for at least two more years, beginning 2009

About the Pandektis project

- 11 archival collections of historical and cultural material
  - Paintings, Monuments, Inscriptions, Portraits, Illustrations, Manuscripts, Press, …
- Material is the output of research by the humanities institutes of the National Hellenic Research Foundation
- > 35000 metadata records
- > 20000 digital assets
- Web repository built using DSpace
  - Live (beta version) since Dec 2007 (http://pandektis.ekt.gr)
- Project co-financed by the European Union and the Greek State under the auspices of the EU Community Support Framework 2000-2006.

Technical aspects of implementation

- Why DSpace?
- Metadata ingest
- Context-sensitive layout
- Linking
- Browsing
- Advanced search

Why DSpace?

- Seriously considered also Fez/Fedora
- Reasons for selecting DSpace:
  - Stable, out-of-the box solution
  - Well documented and architected, relatively easy to extend
  - Team expertise with Java technologies
- Issues of concern related to DSpace as identified in the evaluation phase (referring to Dspace 1.4.2)
  - Support for heterogeneous collections with arbitrary metadata schemas (e.g., in terms of data entry forms, search functionality)
  - Linking between resources.

Metadata representation

- Metadata was available from electronic records created over the years by the researchers
- How did this metadata look like?
  - Mostly non-standard schemata (e.g., custom relational databases, tabular data in Excel)
  - Various formats and tools (Access, Excel, Filemaker, etc.)
  - Duplicate records, typing errors, inconsistent entity naming, metadata available mostly in both Greek and English but correspondence among records not available
- Definition of a common schema for all collections (Dublin Core Application Profile with 29 fields)
  - Each record may contain both Greek and English metadata values

Metadata ingest

- Data pre-processing (de-duplication, linkage, cleaning)
- Mapping of each collection to the common schema.
- Creation of material ready for DSpace ingest script.
- Built logic on an existing, in-house developed, object-oriented frameworks to avoid duplicate code for common tasks (reading from sources, pre-processing, producing output).

Context-sensitive presentation

- What needed to be context-sensitive?
  - Community/collection home page
  - Search/browse results list
  - Record detail view page
  - Presentation of a particular field
- Multi-level adaptive presentation required
  - Per community/collection, per item, per field

Context-sensitive presentation

- Criteria for adaptation
  - Community/collection
  - Type of material (e.g., image/no image)
  - Particular characteristics of the item/digital asset
  - Copyright
  - User profile (language, identity)

Context-sensitive presentation

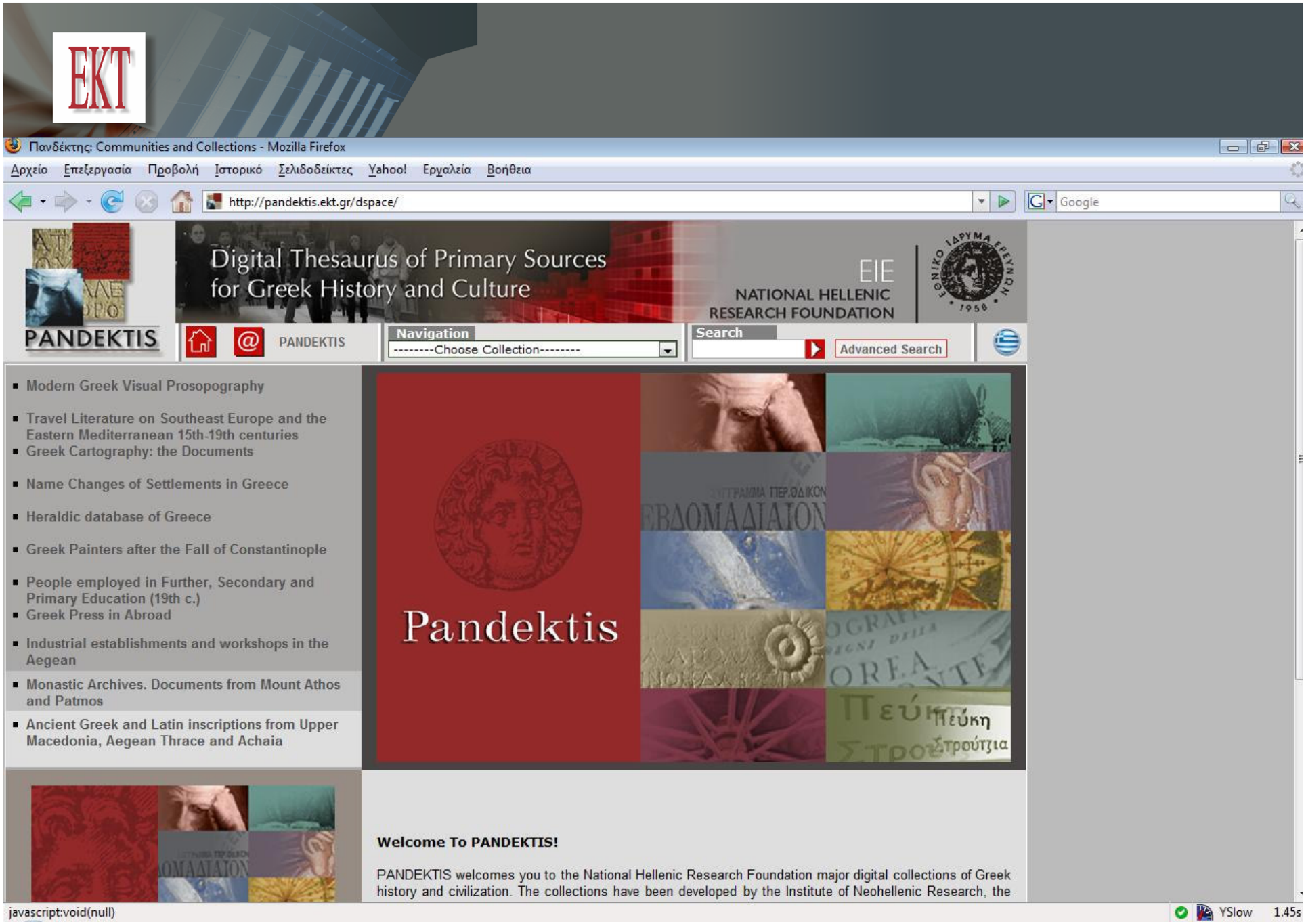
- Light-weight object-oriented framework
  - Models workflow for gathering context, deciding on how to present something and actually presenting it.
- Tasks that are abstracted out and made dynamically interchangeable
  - Context collection
  - Decision logic
  - Rendering of content
- Parameterisation through dependency injection mechanisms

Context-sensitive presentation

- Independent of (and compatible with)
  - UI rendering mechanism (e.g., JSP, Manakin)
  - Repository platform
- Future extension
  - Decision rule specification through a rule engine
  - Addition of actuators (renderers), adaptors, decision functions, context collectors, …

Context-sensitive presentation – example pages from Pandektis

- Screenshots of pages
  1. Pandektis home page
  2. Browse page in image collection
  3. Detailed view of an item (including full-size image)
  4. Detailed record of an item without a digital asset
  5. Detailed view of an item (including only thumbnail-size image due to copyright restrictions)
- In all cases, a collection-specific banner is used for collection branding


EKT
Πανδέκτης: Communities and Collections - Mozilla Firefox
Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Yahoo! Εργαλεία Βοήθεια
http://pandektis.ekt.gr/dspace/
Google
Digital Thesaurus of Primary Sources for Greek History and Culture
EIE
NATIONAL HELLENIC RESEARCH FOUNDATION
ΕΘΝΙΚΟ ΙΔΡΥΜΑ ΕΡΕΥΝΩΝ 1958
PANDEKTIS
PANDEKTIS
Navigation
--------Choose Collection--------
Search
Advanced Search
Modern Greek Visual Prosopography
Travel Literature on Southeast Europe and the Eastern Mediterranean 15th-19th centuries
Greek Cartography: the Documents
Name Changes of Settlements in Greece
Heraldic database of Greece
Greek Painters after the Fall of Constantinople
People employed in Further, Secondary and Primary Education (19th c.)
Greek Press in Abroad
Industrial establishments and workshops in the Aegean
Monastic Archives. Documents from Mount Athos and Patmos
Ancient Greek and Latin inscriptions from Upper Macedonia, Aegean Thrace and Achaia
Pandektis
Welcome To PANDEKTIS!
PANDEKTIS welcomes you to the National Hellenic Research Foundation major digital collections of Greek history and civilization. The collections have been developed by the Institute of Neohellenic Research, the
javascript:void(null)
YSlow 1.45s

Browse by Title

Total Titles : 323 Next page

Showing items 1-21 of 323.

Coat-of-arms of the Abagazi family

Coat-of-arms of the Agalliano family

Coat-of-arms of the Alisafi family

Coat-of-arms of the Alisafi family

Coat-of-arms of the Ammiragli family

Coat-of-arms of the Argyropoulos family

Coat-of-arms of the Barbarigo family

Coat-of-arms of the Callergi family (Calergi)

Πανδέκτης: IThrAeg E011: Ψήφισμα τιμητικό των Αβδηριτών για άγνωστο δανειστή - Mozilla Firefox

Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Yahoo! Εργαλεία Βοήθεια

http://pandektis.ekt.gr/dspace/handle/123456789/163180

Google

ΠΑΝΔΕΚΤΗΣ

ΙΕΡΑ ΙΝΣΤΙΤΟΥΤΟ ΕΛΛΗΝΙΚΗΣ ΚΑΙ ΡΩΜΑΙΚΗΣ ΑΡΧΑΙΟΤΗΤΟΣ

Αρχαίες ελληνικές και λατινικές επιγραφές
Άνω Μακεδονίας, Αιγαιακής Θράκης και Αχαΐας

ΕΘΝΙΚΟ ΙΔΡΥΜΑ ΕΡΕΥΝΩΝ 1958

Ο ΠΑΝΔΕΚΤΗΣ | Περιήγηση ------Επιλέξτε Συλλογή------ | Αναζήτηση | Σύνθετη Αναζήτηση

IThrAeg E011: Ψήφισμα τιμητικό των Αβδηριτών για άγνωστο δανειστή

Πλήρης Εμφάνιση

Τίτλος:
IThrAeg E011: Ψήφισμα τιμητικό των Αβδηριτών για άγνωστο δανειστή

Εναλλακτικός Τίτλος:
IThrAeg E011: Honorific decree of Abdera for an unknown lender

Ημερομηνία:
Τέλη 2ου - αρχές 1ου π.Χ. αι.

Άλλα αναγνωριστικά:
Μουσείο Αβδήρων ΜΑ 5905

Περιγραφή:
Το αριστερό μέρος στήλης από υπόλευκο χονδρόκοκκο μάρμαρο, αποκεκ
άκρο. Μικρές αποκρούσεις κατά τον αριστερό κρόταφο καθιστούν δυσδιά
τα πρώτα γράμματα των στίχων.

Διαστάσεις: 0,34x0,19x0,09.

Ύψος γραμμάτων: 0,01.

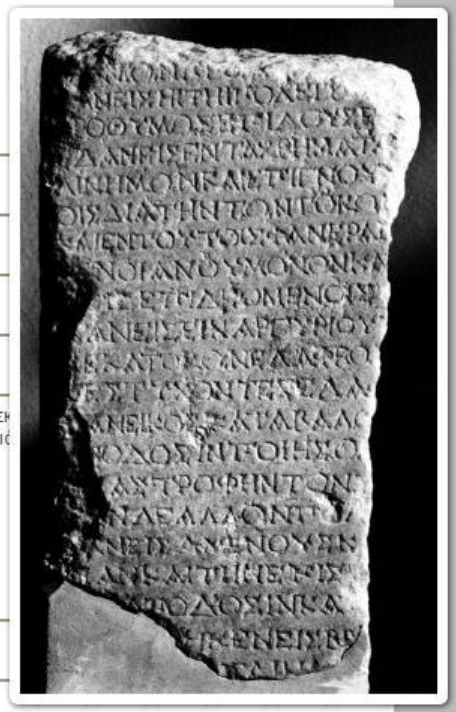
Διάστιχο: στ. 1-8: 0,005, στ. 9-22: 0,007.

μάρμαρο, υπόλευκο, χονδρόκοκκο

Τύπος :
στήλη

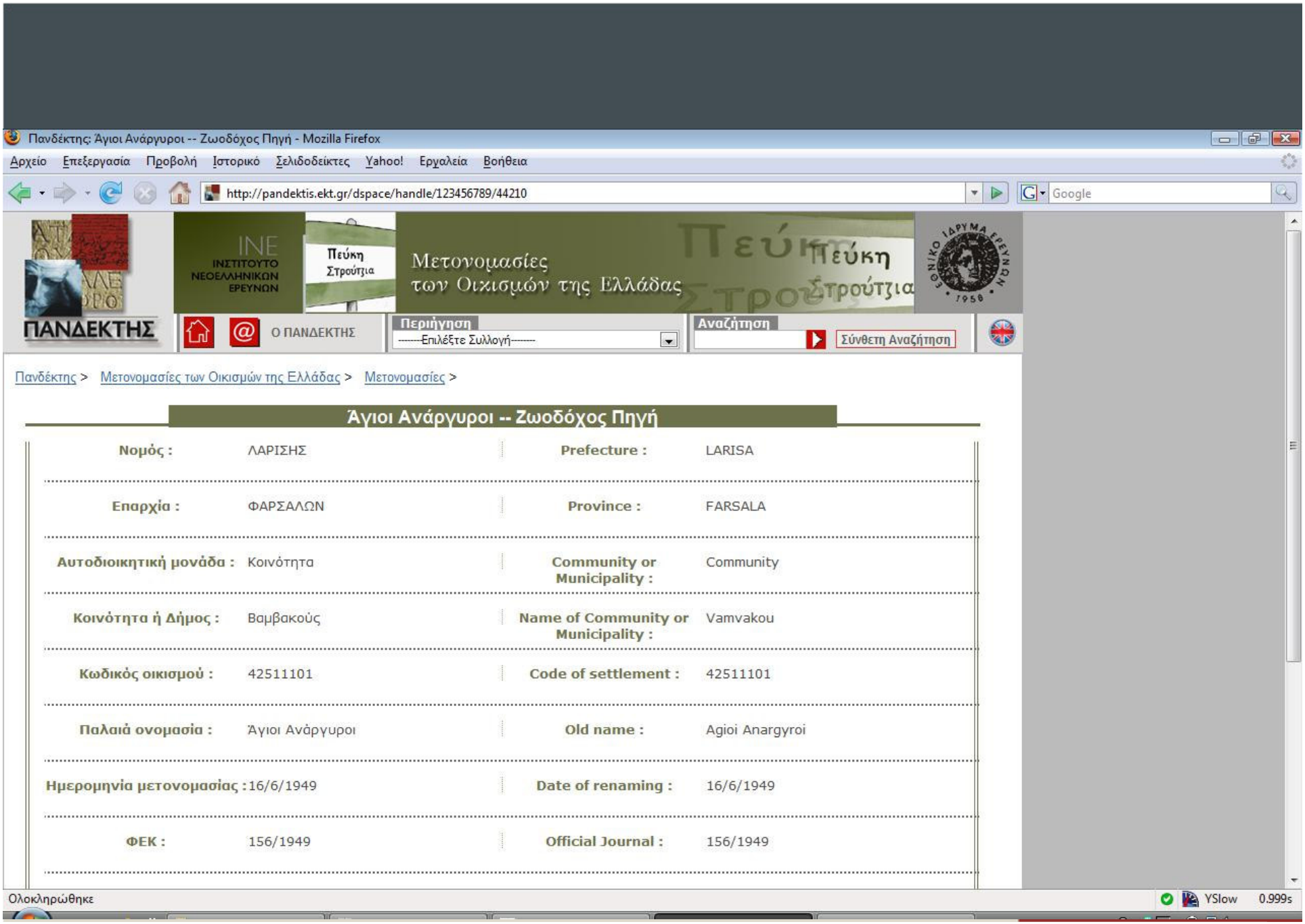
Αρχείο Επιγραφής:
Thraki_E011.html

Τίτλος Έρευνας:



Ολοκληρώθηκε

YSlow 8.374s


Πανδέκτης: Άγιοι Ανάργυροι -- Ζωοδόχος Πηγή - Mozilla Firefox
Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Yahoo! Εργαλεία Βοήθεια
http://pandektis.ekt.gr/dspace/handle/123456789/44210
Google
ΙΝΕ ΙΝΣΤΙΤΟΥΤΟ ΝΕΟΕΛΛΗΝΙΚΩΝ ΕΡΕΥΝΩΝ
Πεύκη Στρούτζια
Μετονομασίες των Οικισμών της Ελλάδας
ΕΘΝΙΚΟ ΙΔΡΥΜΑ ΕΡΕΥΝΩΝ 1958
ΠΑΝΔΕΚΤΗΣ
Ο ΠΑΝΔΕΚΤΗΣ
Περιήγηση
-------Επιλέξτε Συλλογή--------
Αναζήτηση
Σύνθετη Αναζήτηση
Πανδέκτης > Μετονομασίες των Οικισμών της Ελλάδας > Μετονομασίες >
Άγιοι Ανάργυροι -- Ζωοδόχος Πηγή
Νομός : ΛΑΡΙΣΗΣ
Prefecture : LARISA
Επαρχία : ΦΑΡΣΑΛΩΝ
Province : FARSALA
Αυτοδιοικητική μονάδα : Κοινότητα
Community or Municipality : Community
Κοινότητα ή Δήμος : Βαμβακούς
Name of Community or Municipality : Vamvakou
Κωδικός οικισμού : 42511101
Code of settlement : 42511101
Παλαιά ονομασία : Άγιοι Ανάργυροι
Old name : Agioi Anargyroi
Ημερομηνία μετονομασίας : 16/6/1949
Date of renaming : 16/6/1949
ΦΕΚ : 156/1949
Official Journal : 156/1949
Ολοκληρώθηκε
YSlow 0.999s

Πανδέκτης: [Πορτολάνος της Μεσογείου και η δυτική ακτή της Ευρώπης], Georgio Sideri dictus Calapoda cretensis composuit nel anno 1560 - Mozilla Firefox

Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Yahoo! Εργαλεία Βοήθεια

http://pandektis.ekt.gr/dspace/handle/123456789/163327

Google

INR INSTITUTE FOR NEOHELLENIC RESEARCH

Greek Cartography: the Documents (15th-19th centuries)

ΕΘΝΙΚΟ ΙΔΡΥΜΑ ΕΡΕΥΝΩΝ 1958

PANDEKTIS

PANDEKTIS

Navigation

-------Choose Collection--------

Search

Advanced Search

Πανδέκτης > Τεκμήρια Ελληνικής Χαρτογραφίας > Χαρτογραφίες >

Ελληνικά

[Portolan chart of the Mediterranean and the west coast of Europe], Georgio Sideri dictus Calapoda cretensis composuit nel anno 1560

Creator:

Georgio Sideri (Callapoda)

Title:

[Portolan chart of the Mediterranean and the west coast of Europe], Georgio Sideri dictus Calapoda cretensis composuit nel anno 1560

Spatial:

Mediterranean Sea and Atlantic coast of Europe

Date:

1560, 0

Description:

Technique: Manuscript

Category: Portolan Chart

Dimensions: 980 x 570 mm. Provenance: The chart was bought by a Scottish collector in 1958; in the previous year, it had been auctioned by Maggs of London. On the reverse of the chart is the number 24992 or 24972.

Identifiers:

Εδιμβούργο, National Library of Scotland, Ms 20995.

Edinburgh, National Library of Scotland, Ms 20995.

Is Referenced By:

Εύρεση: Επόμενο Προηγούμενο Επισήμανση Ταίριασμα χαρακτήρα

Linking

- Fields values can be links to other resources
  - Internal (other DSpace records)
  - External
- Reciprocal internal linking
  - Books / illustrations
  - Painting / painter
  - Whole / part
- DSpace extension for Pandektis: configurable linking
  - At the DSpace installation level
  - At the context-sensitive presentation level

Other issues

- Browsing
  - Multi-lingual field values (e.g., titles in Greek / English) => significant adjustments have been required
  - Scalability / Configurability
  - Some of the browsing issues hopefully obsolete with DSpace 1.5
- Configurable i18n
  - Requirement: user is able to change the language in every page
  - Implemented according to the philosophy of an older DSpace patch

Advanced search – range queries

- We had an explicit requirement for the support of range queries in advanced search
  - "Get me all works by painter XXX dated between 1650 and 1670"
- The Lucene search engine supports this, however needed to do some adjustments to make it work with DSpace 1.4.2
  - Needed to modify the DSpace code that gets the Lucece query from the advanced form and passes it over to Lucene
  - Enhanced the advanced search form with some additional XHTML/Javascript code to enable the user to specify range queries

Deployment

- Three tier system:
  - Redundancy in every tier
  - Web tier: CentOS Cluster
  - Database tier: Oracle Real Application Cluster 10g
  - Application tier: XEN VM
    - run over a VM host platform of four Opteron-based servers sharing resources with other production VMs
    - two server configurations with 24 cores and 40GB of RAM in total
  - Overall infrastructure
    - Pandektis shares space from two SANs with 44TB Storage space in in total.
      - Expansion is underway.
    - Two redundant tape libraries
    - Gbps connection to GRNET/GEANT
      - 10 fold increase feasible
    - Nagios system monitoring in place

Further work

- Faceted browsing
  - Dwell?
- Browse by map
  - Flash / Google maps mashups
  - Requires high accuracy data cleaning
- Data entry for heterogeneous metadata schemas
- Specialised tools for intelligent data processing
  - Data cleaning/entity name resolution/record linkage
  - Automated extraction of descriptive metadata
- Improving multi-lingual support
  - Search (e.g., stress marks), browsing (DSpace 1.5?)

More info

nhoussos AT ekt.gr

http://pandektis.ekt.gr

http://openaccess.gr