A service-oriented national e-thesis information system and repository

Nikos Houssos

Panagiotis Stathopoulos
Ioanna Sarantopoulou
Dimitris Zavaliadis
Evi Sachini

National Documentation Centre (EKT)



Agenda

- The Hellenic National Archive of Doctoral Dissertations (HEDI) - Introduction
- System overview and architecture
- Design choices and technical challenges
- Repositories interoperating in a SOA environment



Hellenic National Archive of PhD Theses

- By law all theses awarded by Greek Universities
- Theses of Greek scholars for PhDs obtained in foreign universities
- In operation at EKT (print archive) since 1985: submission of theses by universities to EKT, centralised cataloguing in UNIMARC
- Open online access to metadata and full-text since late 90's – Z39.50 compliant bibliographic system (ARGO - http://argo.ekt.gr)
- 24000 theses in total (since 1901)
- 25% theses not yet digitized
- 1200-1400 arriving every year

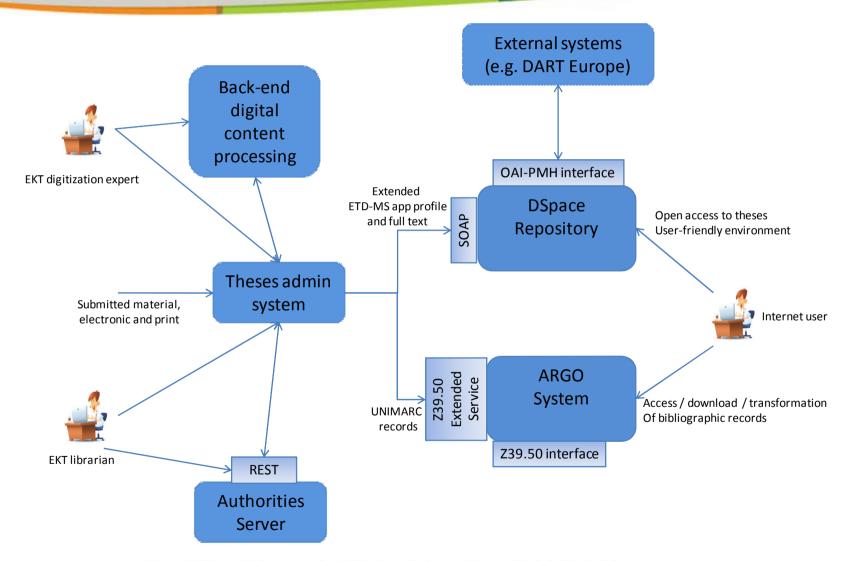


HEDI – Goals for new version

- Make theses available through DSpace
- Develop administration application to support and monitor workflows for processing incoming material
- Create separate authority servers
- Migrate to an open source software infrastructure from operating system to repository software platform
- Make the whole system work with individual components in a SOA configuration



System architecture and overview





The administration application functions

- Record incoming material and monitor its status throughout its lifecycle – both for print and electronic material
- Support and monitor internal workflows for processing digital content
- Reporting on aspects like throughput of internal workflows, numbers of material in each processing state, etc.
- Cataloguing and export in UNIMARC
- Production of appropriate UNIMARC and DC records and update of both ARGO and DSpace repository

Benefits of introducing DSpace

- User friendly interface, much better browsing facilities
- Easier compatibility with aggregators / harvesters (OAI-PMH), persistent identifier systems, search engines, web 2.0 systems, etc.



The DSpace e-thesis repository

- Available at http://phdtheses.ekt.gr
- Contributor to DART Europe via OAI-PMH harvesting
- Main customisations:
 - Web services for CRUD operations to metadata records
 - Service for uploading digital material to DSpace
 - Custom application profile based on ETD-MS
 - Frascati-based classification based initially on department / lab / research group information
 - Enhance search to support stress independent search
 - Various UI customisations



Implementation technologies

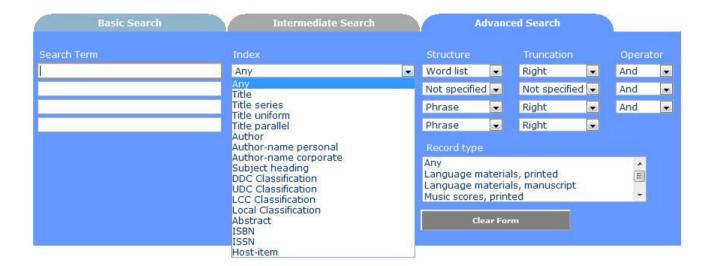
- Groovy / Grails great web applications framework, great integration with Java and Spring
- Lightweight workflow management engine built inhouse
 - Models workflows as finite state machines
 - Uses Spring dependency injection for configuration
- SOAP and REST web services
- Ruby implementation of the Z39.50 Database Update Extended used to update ARGO
- Custom solution based on Spring framework to upload digital file(s) to DSpace



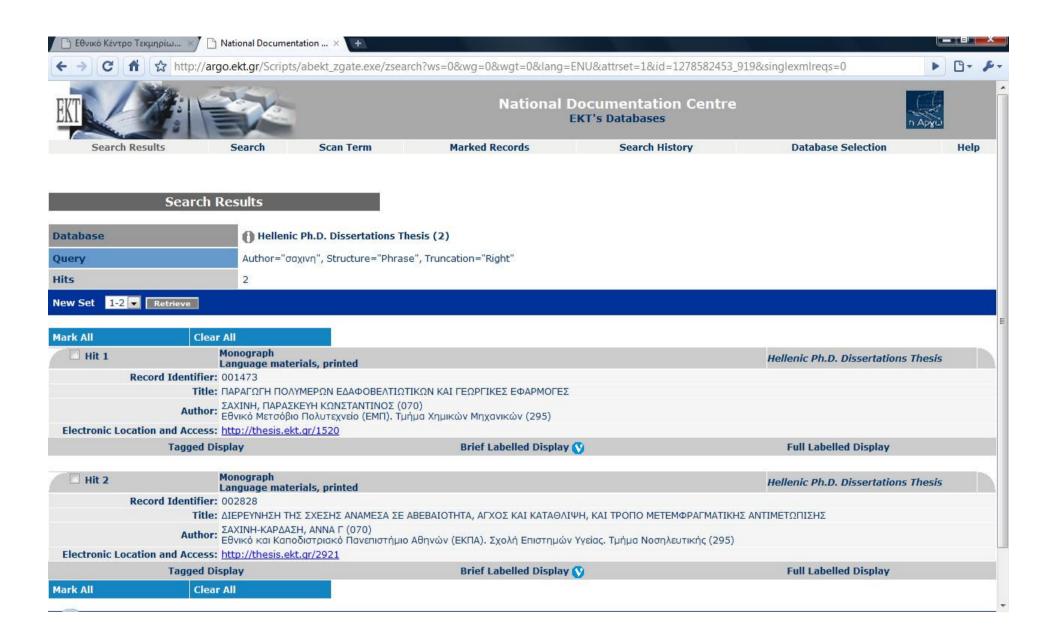


Search Hellenic Ph.D. Dissertations Thesis

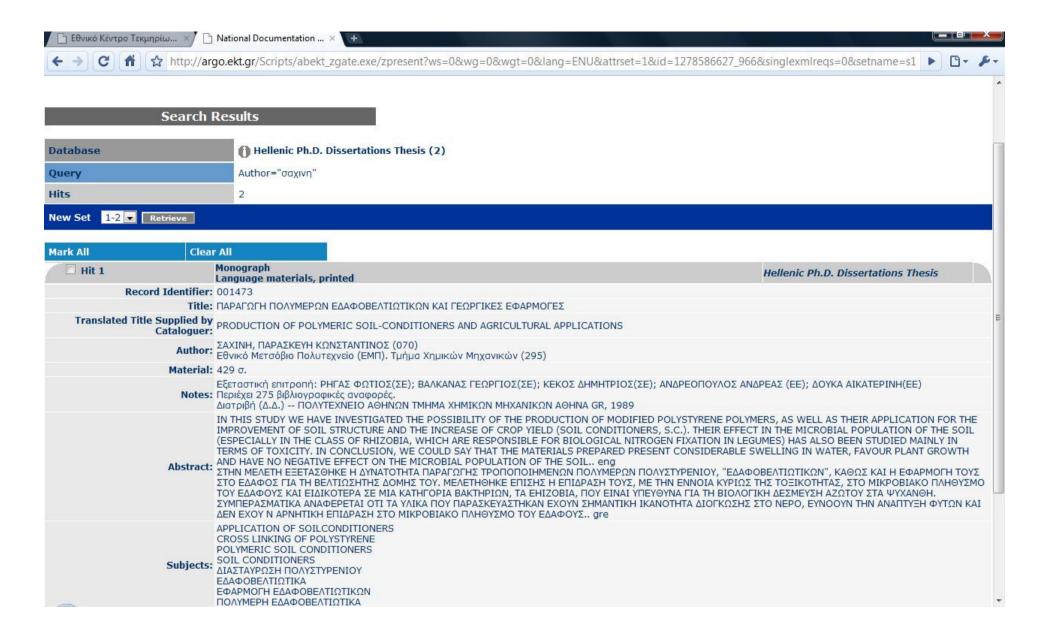
Database













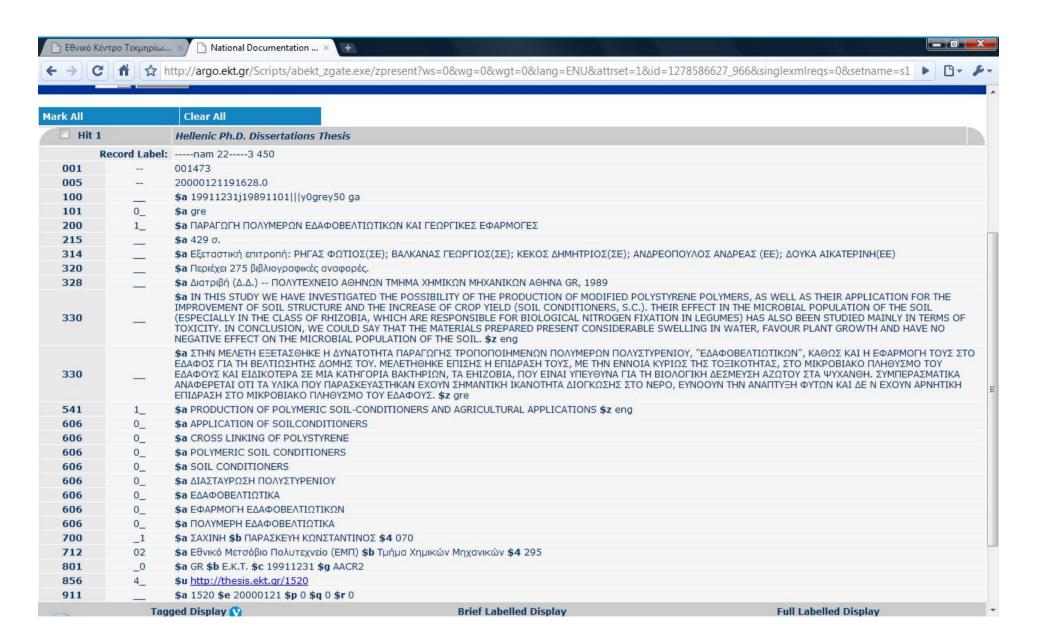


Saving records in ISO2709 or MARCXML format

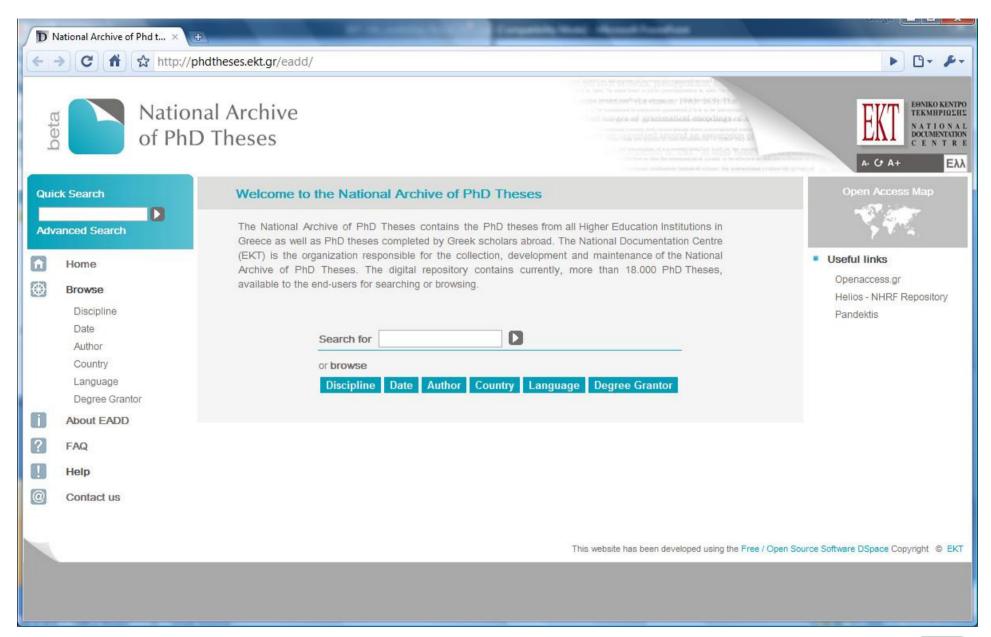
Database	Hellenic Ph.D. Dissertations Thesis
Marked Records	1

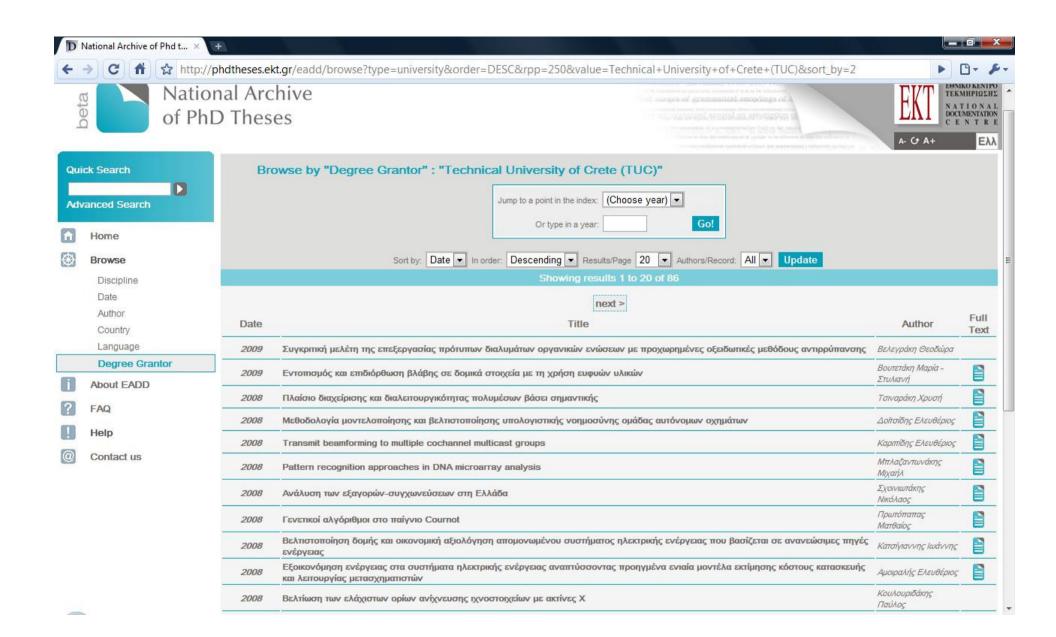
■ ISO2709 ■ MARCXMI	-
2. Select the saving cha	racter set of marked records
O ANSEL	Windows:Western
© ISO 5428	Windows:Central Europe
© Elot 928	Windows:Greek
© UTF-8	Windows:Cyrillic
Save	Return



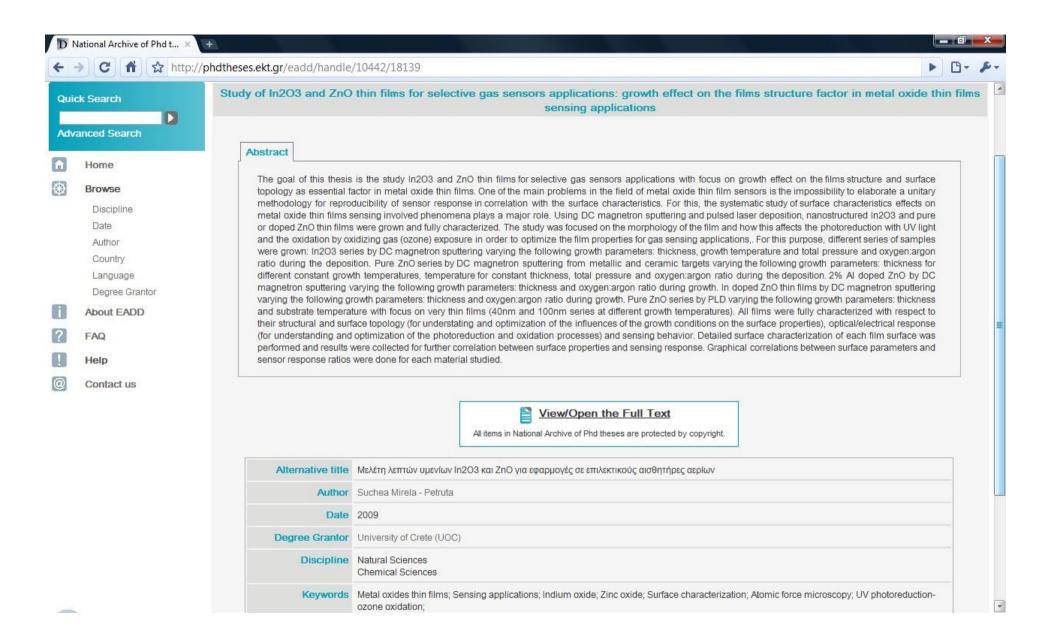




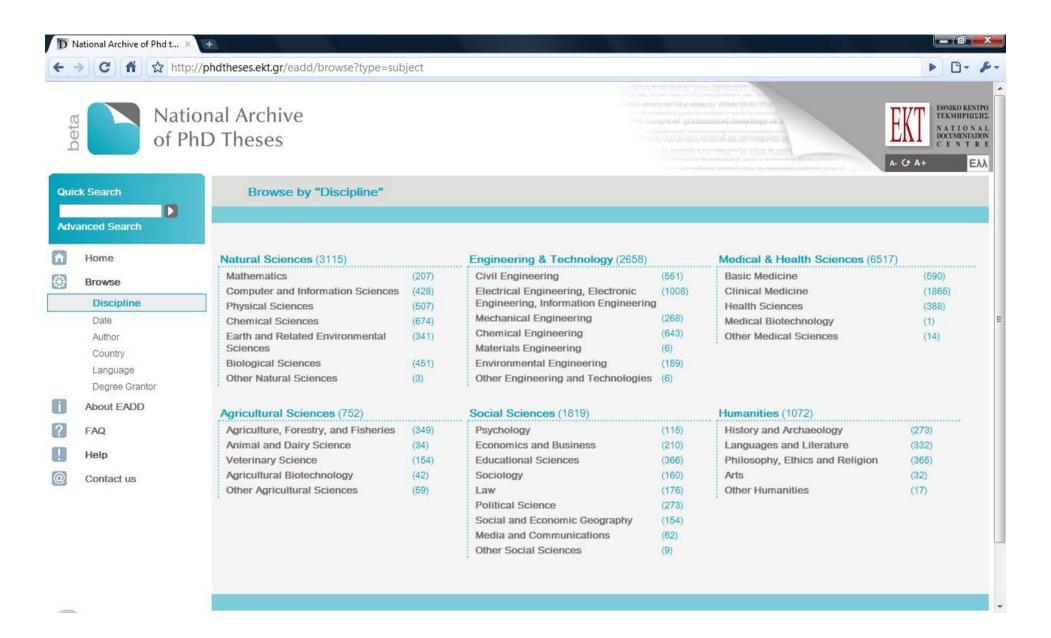




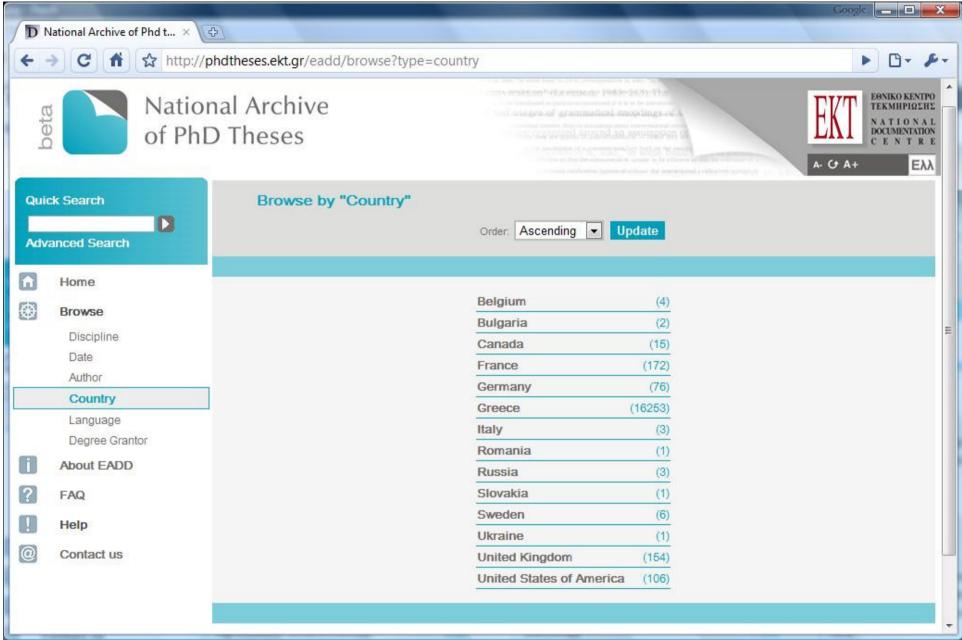








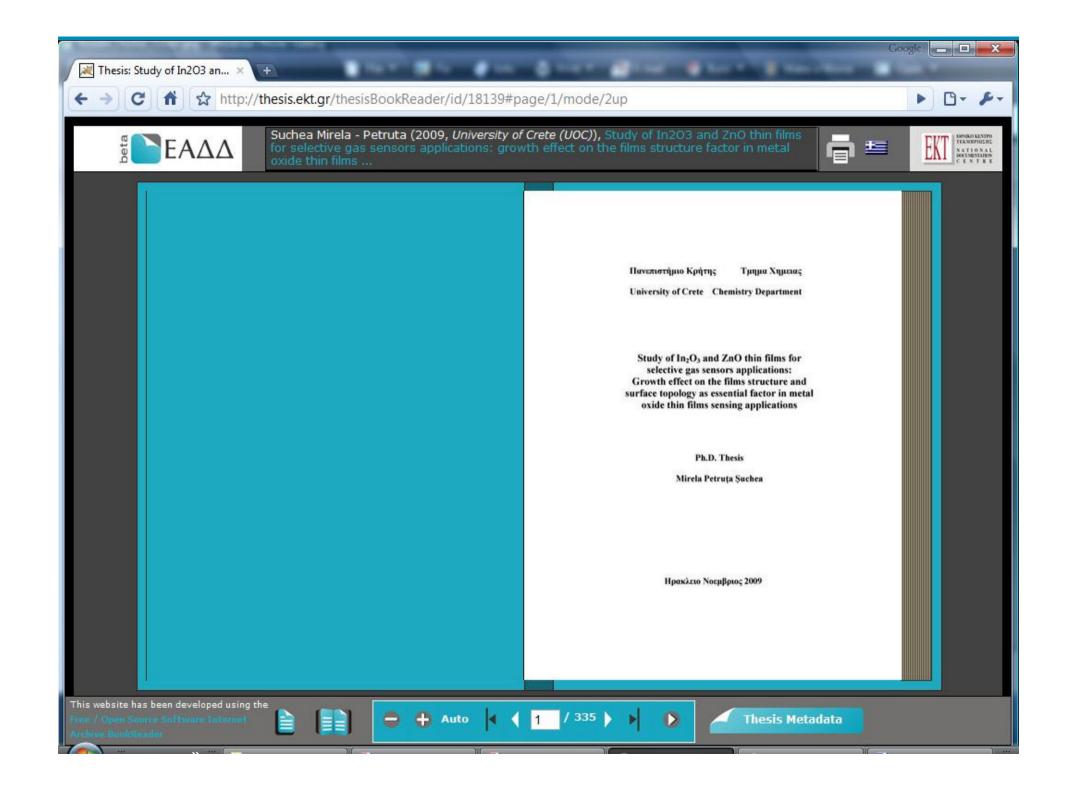




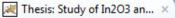
Online reading

- An alternative way to present content opening a range of possibilities
- Better promotion of the overall system back links to repository
- Put burden on digital content handling
- Using Internet Archive Book Reader
- On-going work
 - Full-text search with hit highlighting
 - Bookmarking
 - Detailed access statistics navigation / bookmarking

















http://thesis.ekt.gr/thesisBookReader/id/18139#page/44/mode/2up









Suchea Mirela - Petruta (2009, University of Crete (UOC)), Study of In2O3 and ZnO thin films





- . Thermionic and field emission between bulk and surface electronic states with possible reflections at the surface
- · Generation and recombination in the bulk
- · Both majority and minority carrier transport in the bulk
- Thermionic and field emission between electrode contacts and the bulk.

Furthermore, the exact surface electronic structure is often not known. In the case of strong inversion, the analysis is quite analogous to that used in semiconductor device physics. Even if the coupling is well described, the calculation is not simple and the system can probably be simulated only numerically.

The analysis of changes in charge carrier mobility near the surface is often very complicated. The effect of surface roughness on the effective mobility of surface excess carrier due to band bending near the surface has been evaluated by Greene et al. [12]. A simpler analysis, giving out almost the same expressions, valid only for depletion and strong accumulation, can be found in reference 13. Later, more simple approximate expressions have been derived by Goldstein et al. [14]. For scattering by fixed surface charges under strong inversion it may be also possible to adopt an approach developed for MOSFET transistors [15].

If no Fermi level pinning is present at the surface, the screening length is given by the bulk extrinsic Debye length [13], given by:

$$L_{D} = \left(\frac{e \varepsilon_{0} k_{B} T}{q^{2} (p+n)}\right)^{1}$$

where ε and ε₀ are the specific and vacuum permittivity, k₈ is Boltzmann constant, T is the temperature, q is the electron charge and n and p are the densities of free electrons and holes respectively. In case of Fermi level pinning by "metal like" surface states, the scattering potential in the bulk will be a quadrupole potential decreasing as 1/r3. The screening in the surface plane is given by the 2D screening length [16] given by:

$$\beta^{-1} = \frac{1}{2} \sqrt{a_s d}$$

where a_B = 4πε_{Co}h²/m*q² is the Bohr radius, m* being the effective mass and d is the thickness of the 2D system. The surface screening length is of the order of a

All this information is very useful and applies guite well locally in a very narrow region or for epitaxial growth, but it becomes practically impossible to be used for the explanation/understanding of the behavior of real polycrystalline films.

Conduction mechanisms

The oldest models, regarding conduction mechanism in polycrystalline films, are based on the grain boundary conduction model - developed by Petritz [17] in 1956, which is based on the assumption that the conductivity behavior in polycrystalline films closely approaches that of semiconductors with predominant grain boundary conduction mechanism. The carrier mobility in these films is limited by scattering at the surface and the grain boundaries as well as by normal bulk processes. A model of intergrain boundaries affected by the diffusion of an active gas has been used by Seager and Ginley [18] to explain the changes of conductivity seen in polycrystalline silicon. As found out, diffusion of oxygen down the grain boundaries promoted in these regions significant changes in the density of defect states, resulting in a decrease of conductivity. This model, described extensively by

Seager and Castner for the case of polycrystalline silicon [19], has been adopted until now as the basic approach to explain the conductivity mechanism in polycrystalline metal oxide films. The validity of this model has been confirmed in polycrystalline metal oxide films by experimental results related to the dependence of the conductivity on the temperature, but it cannot explain experimental results when the films are used as sensing layers or when photoreduction is involved.

The main features of this conduction model are: conduction from grain to grain, disturbed by surface barriers which are strongly influenced by chemisorbed

The formation of potential barriers at the grain boundaries was proposed by Petritz. [17] in 1956, in addition to the normal lattice discontinuity caused by the boundaries. Other models have also been proposed to explain the transport behavior due to the grain boundaries, as that of Volger [20] and Berger [21]. Since then, this subject has been reviewed in detail by Kazmerski [22], while, more trials also appeared using different approaches like the ones of Gardner [23, 24]. Bårsan and Weimar [25, 26].

Petritz theory constitutes the basic theoretical analysis of transport mechanisms in polycrystalline semiconducting films. According to this model, grain boundary potential barriers are formed in an n-type semiconductor when the grain boundary region has a lower chemical potential (Fermi level, Ex) for majority carriers, than the grains, due to the density of defect states in this region. These defect states can appear due to the tendency of grain boundaries to act as diffusion whirlpool for impurities. Therefore, these defect states can be treated as trapping centers for majority carriers, resulting in a reduction of their concentration in the boundary region. This in turn causes a flux of majority carriers into the boundary region. creating a space charge build up at these boundaries, which prevents further flux of majority carriers and therefore forms a depletion region for them. This can be presented in a band diagram by an upward bending of the conduction and valence band edges. For a p-type semiconductor respectively, the band edges bend down. toward the Fermi level

The accumulated negative charge near a joint force the energy bands to be bend upwards by an amount of Φ_8 . Since the Fermi energy at equilibrium must be continuous over the grain boundary, the height of the potential barrier, eQs, will be given by the difference of the Fermi's grain-boundary energies. Majority carriers can cross over a grain boundary potential barrier, following two different mechanisms. One is the thermal emission over the barrier and the other is the quantum mechanical tunneling. For the evaluation of electrical characteristics of semiconducting films, most models compare the behavior of the films to that of the bulk crystal. If the bulk crystal was perfect, the conduction carriers could flow unimpeded in a perfect periodic potential. In a real bulk crystal lattice, vibrations, impurities and defects can cause deviations from the ideal behavior, an approach that can be used in polycrystalline thin films analysis, which, however, can result is quite inexact results. The carrier mobility is related directly to the mean free time between collisions, which in turn is determined by the various scattering mechanisms. For bulk crystal behavior in semiconductors, two scattering processes are important: lattice scattering and ionized impurity scattering. In polycrystalline semiconducting films, however, the effect of the grain boundaries should be also considered as an additional scattering mechanism for the carriers. The carriers collide at the grain boundaries and, in a steady state, have an effective mean free path λ_{G} , constrained by the size of the grains, and a mean









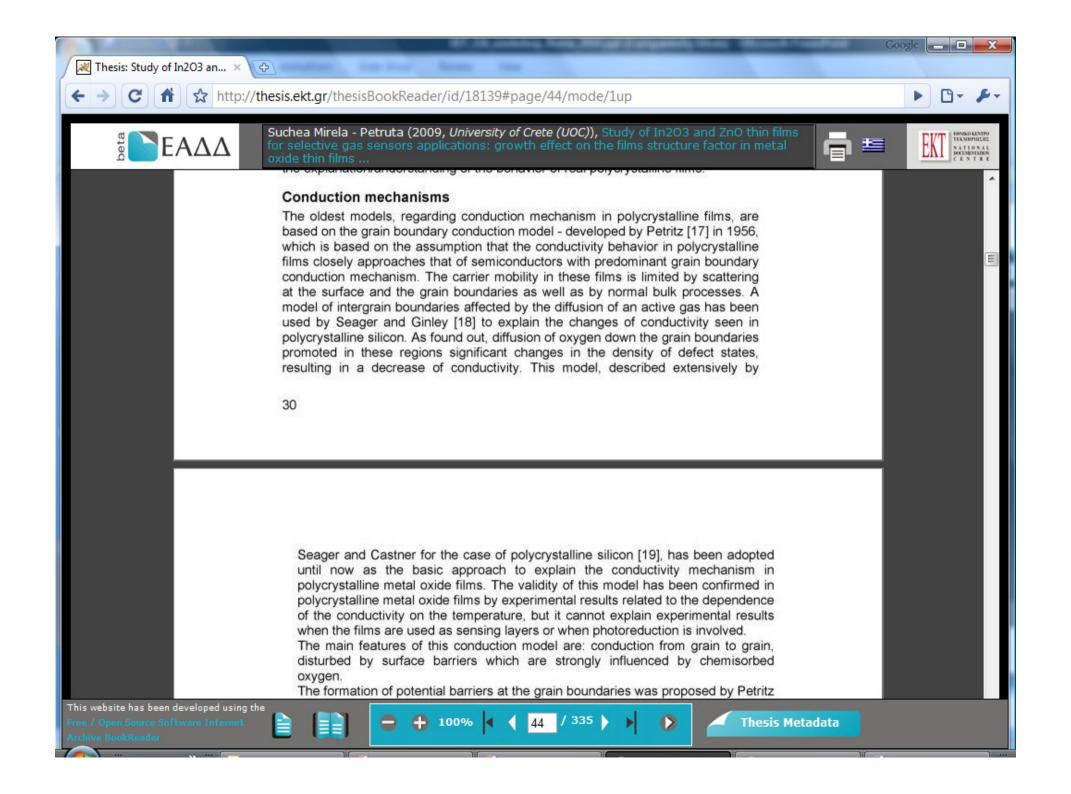












Design choices and technical challenges (1)

- Adoption of a SOA approach with DSpace being an crucial distinct part of the overall architecture
- Continuation of UNIMARC cataloguing support of the ARGO bibliographic portal
- Adoption of software stacks consisting of open source or home grown components



Design choices and technical challenges (2)

- Support of distributed transactions
 - System should gracefully recover from update failures
 - Ad hoc implementation of transaction behaviour
- Cleaning and enhancing metadata records for theses
- Data migration



The underlying infrastructure

- Gradual migration of modules in parallel with the software development process
- From closed source software, & proprietary hardware to a fully open source stack using a virtualisation platform over commodity hardware.
 - From Filenet, Oracle 9, and 2 proprietary Operating Systems
 - DSpace, Postgress, and the CentOS distribution
 - End result: Fully open source, fully virtualised.
- Virtualisation: computing and storage resources allocated, are flexible and dynamic.
 - Includes each VM memory, number of processors, disk space, without reboot for many of them
 - Resource pool of up to 48 processors and 160GB memory to be allocated for the servers comprising the system, TBs of disk space



Resources, monitoring and mgt

- Resources allocated based on observed performance
 - No overdimensioning, need to have resources when needed
 - "monitor and control" loop (broadly basde on IT Service Lifecycle definition)
- Monitoring system for the full range of each of the platform elements, tiers, software modules in all levels:
 - From hosts to NFS fileshares and DBs and random's pages content sensitive monitoring
 - SOA puts stress on that, large number of elements that must be monitored
 - Virtualization: added value, "Green" Hedi system, can check at http://code.google.com/p/e-vigr/ (OK it's the whole platform not just HEDI!)



Further work

- Application of automated metadata extraction from full-text theses to assist cataloguing
- Integration of digital content back-end processing workflows – full migration from a commercial to an open source environment
- Automation of the procedure of digital file quality checking
- Interconnection with CRIS systems linking theses with authors, organisations and projects which have specifically funded PhD theses



Thank you

- More information:
 - Nikos Houssos / nhoussos AT ekt.gr
 - Panagiotis Stathopoulos / pstath AT ekt.gr

