

# What is a Digital Library?

Modeling things is making our life easier: the Reference Model

Yannis Ioannidis
University of Athens

Theory and practice in Digital Libraries: a European approach Athens, 13 December 2010





# DL.org Network of DL Stakeholders

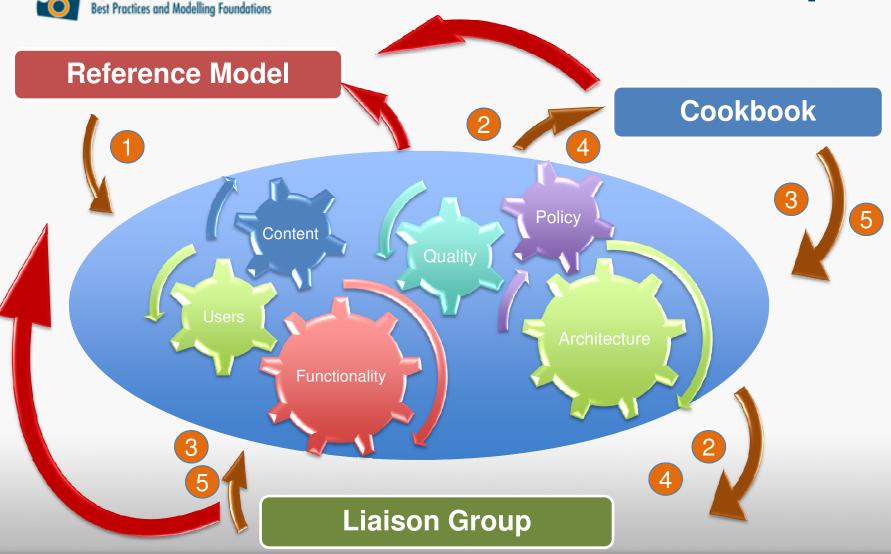
WGs on six Digital Library Universe Domains

[DELOS DL Reference Model]





# **Interactions and Output**



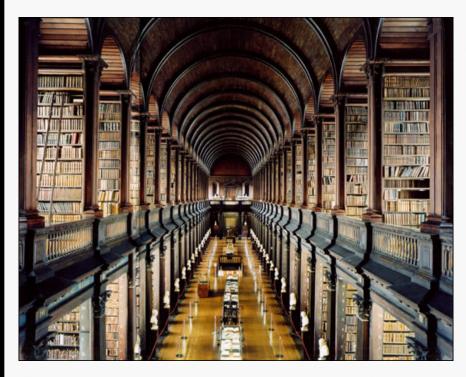


## **Outline**

- Motivations
- RM Overview
- ☐ The RM Domains
- Discussion



# Libraries....

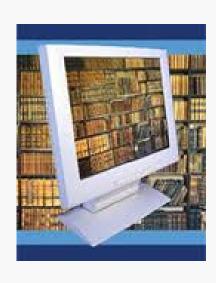




and...



## ...Libraries?

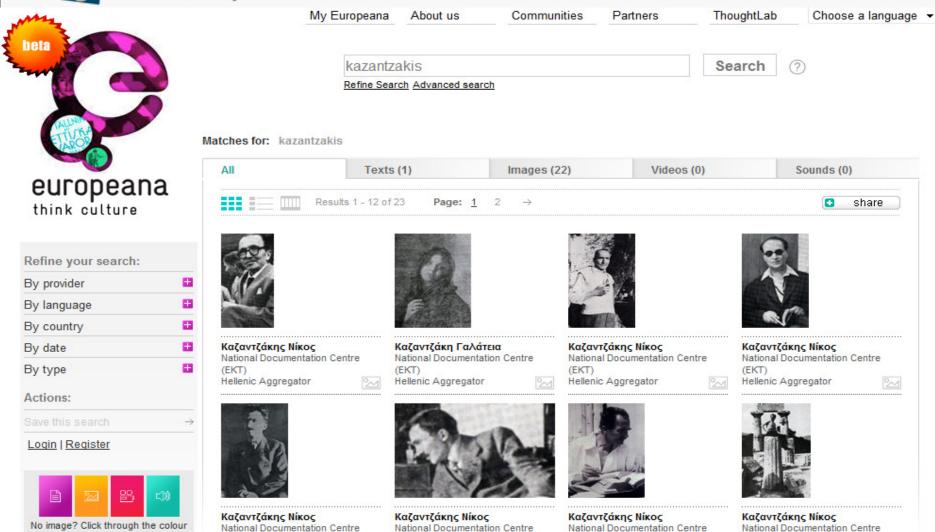








# Europeana



block to see the item. We are still



# **ACM Digital Library**



Timos Sellis	SEARCH

SIGN IN SIGN LIP

Searching for: Timos Sellis (start a new search)

Found 407 of 1,602,563 within The ACM Guide to Computing Literature

Limit your search to Publications from ACM and Affiliated Organizations

REFINE YOUR SEARCH	Search Results Related Journals Related Magazines Related SIGs Related Conferences
Refine by Keywords	Results 1 - 20 of 407 Sort by relevance   ▼ in expanded form ▼
Timos Sellis	Result page: 1 2 3 4 5 6 7 8 9 10 next >>
Discovered Terms	1 The generalized pre-grouping transformation: aggregate-query optimization in the presence of dependencies  Aris Tsois, Timos Sellis  September 2003 VLDB '03: Proceedings of the 29th international conference on Very large data bases - Volume
Refine by People	29 , Volume 29
Names	Publisher: VLDB Endowment
Institutions Authors	Full text available: Pdf (345.15 KB)
Editors Advisors	Bibliometrics: Downloads (6 Weeks): 0, Downloads (12 Months): 20, Downloads (Overall): 69, Citation Count: 2
Reviewers	One of the recently proposed techniques for the efficient evaluation of OLAP aggregate queries is the usage of clustering
Refine by Publications	access methods. These methods store the fact table of a data warehouse clustered according to the dimension hierarchies using special
Publication Year Publication Names	
ACM Publications All Publications	2 <u>Data Warehouse Configuration</u> Dimitri Theodoratos, Timos K. Sellis
Content Formats	August 1997 VLDB '97: Proceedings of the 23rd International Conference on Very Large Data Bases
<u>Publishers</u>	Publisher: Morgan Kaufmann Publishers Inc.
Refine by Conferences Sponsors	Bibliometrics: Downloads (6 Weeks): n/a, Downloads (12 Months): n/a, Downloads (Overall): n/a, Citation Count: 69
Events	
Proceeding Series	3 Data management research at the Knowledge and Database Systems Lab: (NTU Athens)
ADVANCED SEARCH	Timos Sellis, Yannis Vassiliou  June 2006 SIGMOD Record , Volume 35 Issue 2
Advanced Search	Publisher: ACM



## **Amazon**

amazon.com Prime

Hello, Natalia Manola. We have recommendations for you. (Not Natalia?)

Natalia's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards

Search Literature & Fiction Shop All Departments kazantzakis Books Advanced Search **Browse Subjects** New Releases Bestsellers The New York Times® Bestsellers

#### New Releases

#### **Any New Release**

Coming Soon (1)

#### Department

- < Any Department
- < Books

#### Literature & Fiction

Genre Fiction (8)

Contemporary (24)

Literary (8)

History & Criticism (31)

United States (14)

World Literature (6)

Books & Reading (1)

Poetry (6)

Foreign Language

Fiction (12)

Classics (4)

Comic (1)

British (1)

#### Format

#### **Any Format**

HTML (3)

Kindle Books (1)

Printed Books (68)

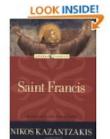
#### Binding

#### Any Binding

Paperback (45)

#### Books > "kazantzakis" > Literature & Fiction

#### Showing 1 - 12 of 82 Results



Saint Francis by Nikos Kazantzakis and John Michael Talbot (Apr 2005)

★★★★ ▼ (9 customer reviews)

Formats	Buy new	New from	Used from Co	ollectible from
Paperback Order in the next 25 hours to get it by Tuesday, Dec 14. Only 2 left in stock - order soon.	\$13.95 \$11.16 <i>Prime</i>	\$9.10	\$8.98	\$7.49
Hardcover			\$7.93	\$45.00

Excerpt - page 7: "... Nikos Kazantzakis 7 All the soil inside me has blossomed, Father Francis. ..." Surprise me! See a random page in this book.





or CHRIST

The Last Temptation of Christ by Nikos Kazantzakis (Mar 1, 1998)

★★★★ ▼ (91 customer reviews)

Formats	Buy new	New from	Used from Co	llectible from
Paperback Order in the next 27 hours to get it by Tuesday, Dec 14.	\$16.00 \$10.88 Prime	\$7.60	\$0.01	\$9.95
Hardcover			\$35.00	\$49.95

Excerpt - Front Cover: "... NIKOS KAZANTZAKIS victo IIIThe Ne York Time Book Be THE LAST TEMPTATION Surprise me! See a random page in this book.

Report to Greco by Nikos Kazantzakis (Jun 18, 2001)

3.



# **Google Scholar**

Google scholar Timos Sellis Search					
Scholar Articles and patents	anytime ▼	include citations	▼	ate email alert	
[PDF] The R-tree: A dynamic index for multi-dimensional objects  T Sellis, N Roussopoulos The VLDB Journal, 1987 - Citeseer The problem of indexing multidimensional objects is considered. First, a classification of existing methods is given along with a discussion of the major issues involved in multidimensional data indexing. Second, a variation to Guttman's R-trees (R + -trees) that avoids overlapping  Cited by 1274 - Related articles - View as HTML - All 35 versions					
Multiple-query optimization TK Sellis - ACM Transactions on Database Systems (TODS), 1988 - portal.acm.org Some recently proposed extensions to relational database systems, as well as to deductive database systems, require support for multiple-query processing. For example, in a database system enhanced with inference capabilities, a simple query involving a rule with multiple definitions may  Cited by 495 - Related articles - All 19 versions					
A model for the prediction of R-tree perf, T Sellis - Proceedings of the fifteenth ACM SIGA Abstract: In this paper we present an analytical mo (and its variants) when a range query needs to be of the dataset only, ie, the proposed formula that excited by 267 - Related articles - BL Direct - All 23 versions.	CT, 1996 - po del that predicts t answered. The co stimates the num	he performance of R-tre ost model uses knowled	dge		[PDF] from psu.edu
Topological relations in the world of minimal D Papadias, T Sellis, Y Theodoridis Proceeding Dimitris Papadias Department of Computer Science California, San Diego CA 92093-0114, USA dimitris of Electrical and Computer Engineering National T Cited by 235 - Related articles - BL Direct - All 23 versions.	gs of the, 1995 ce and Engineerir s@cs.ucsd. edu echnical Universit	- portal.acm.org ng University of Times <b>Sellis</b> Departm	ent	<u>ees</u>	[PDF] from psu.edu



## Google



### kazantzakis

×

Search

About 327,000 results (0.24 seconds)

Advanced search

### Everything

- Images
- Videos
- **Books**
- ▼ More

#### Any time

Past 3 weeks

#### All results

Sites with images

■ More search tools

#### Something different

dostoevsky thucydides flaubert turgenev sophocles

## ► Nikos **Kazantzakis** - Wikipedia, the free encyclopedia ☆ ♀

Nikos Kazantzakis (Greek: Νίκος Καζαντζάκης) (February 18, 1883, Heraklion, Crete, Ottoman Empire - October 26, 1957, Freiburg, Germany) was arguably the ... Biography - Literary work - Bibliography of English ... - Further reading en.wikipedia.org/wiki/Nikos Kazantzakis - Cached - Similar

#### Nikos Kazantzakis 😭 🔍

Nikos **Kazantzakis** was born in Megalokastro, Ottoman Empire, now Iráklion, Crete, the son of Michael **Kazantzakis**, a farmer and dealer of in animal feed, ... kirjasto.sci.fi/kazantza.htm - Cached

#### Nikos Kazantzakis Quotes 😭 🔍

12 quotes and quotations by Nikos **Kazantzakis**. ... Nikos **Kazantzakis** Beauty is merciless. You do not look at it, it looks at you and does not forgive. ... www.brainyquote.com/quotes/.../n/nikos **kazantzakis**.html - Cached - Similar

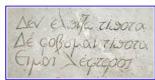
### Nikos Kazantzakis - Wikiquote 😭 🔍

Nikos **Kazantzakis** (1883-02-18 – 1957-10-26) was a Greek novelist, poet, playwright and ... As quoted in Nikos **Kazantzakis** (1968) by Helen **Kazantzakis** ... en.wikiquote.org/wiki/Nikos **Kazantzakis** - Cached - Similar

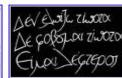
#### Images for kazantzakis - Report images











### The Nikos Kazantzakis Files 🛱 🔍

The Nikos **Kazantzakis** Files first appeared on the World Wide Web in 1997, to mark the fortieth anniversary of the author's death. They were compiled by the ... www.historical-museum.gr/kazantzakis/en/index.html - Cached - Similar



# What is Digital Library?

- Simulation of "real" Library?
- ☐ Digitized/Digital Content?
- Digital Repository?
- Data/Knowledge Base ?
- Webpage?
- ☐ User Online Community
- Online Organization ?
- ☐ Heritage Preservation tool?
- eLearning tool?
- Research tool?

None of these, all of these, and many more!!!



## **DL Universe**

DIENST ACM DL **DSPACE** PERSEUS PAPYRUS OPENDLIB
DASCIENCE **IMPACT FEDORA** OPENAIRE TEI DILIGENT e-FRAMEWORK **BRICKS** 



# **Digital Library Issues**

- Comparison among systems is hard
  - Different focus
  - Different concepts
  - Different terminology
- No guidelines for DL education
- ☐ Lack of DL systems design and development methodologies
- ☐ No systematic approach to interoperability & integration of

solutions

Lack of foundations!



## **Reference Model**

- A reference model is an abstract framework for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment
- A reference model is based on a small number of unifying concepts and may be used as a basis for education and explaining standards to a non-specialist
- A reference model is not directly tied to any standards, technologies or other concrete implementation details, but it does seek to provide a common semantics that can be used unambiguously across and between different implementations

2005 Adobe Systems Incorporated. All Rights Reserved



## **DL Reference Model**

## **Objective**

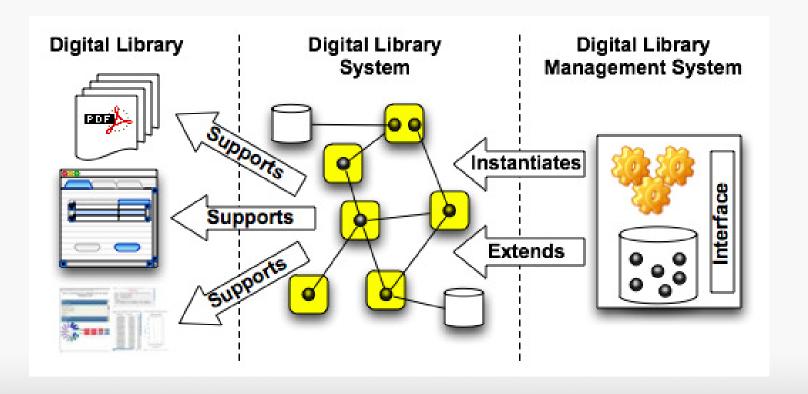
To set the foundations and identify the cornerstone concepts within the universe of Digital Libraries, facilitating the integration of research and proposing better ways of developing appropriate systems

## Consists of 3 parts:

- Digital Library Manifesto
- Digital Library Reference Model in a Nutshell
- Digital Library Reference Model Concepts & Relations.



# The DL "Systems"

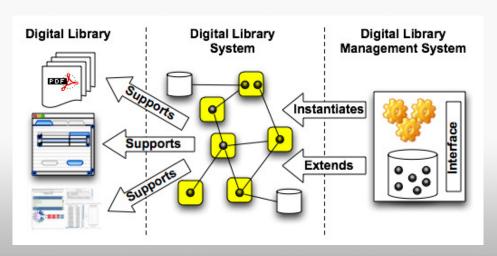




# **Digital Library**

A (potentially virtual) **organization** that comprehensively collects, manages, and preserves for the *long term, rich curated* digital **content**, and offers to its *targeted* **user** communities *specialized* **functionality** on that content, of *measurable high* **quality** and according to *comprehensive codified* **policies**.

The DELOS Digital Library Reference Model





# **Digital Library**

## ABSTRACT CHARACTERISTICS MAIN CONCEPTS

long term organization

rich curated content

targeted users

specialized functionality

measurable high quality

comprehensive codified policies



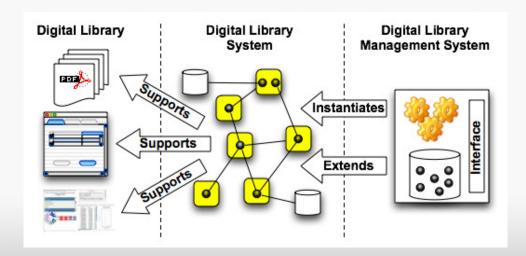
# **Digital Library**

- Concepts for any info environment or system
- Characteristics distinguish DLs
  - Abstract and subject to interpretation
  - Conceptual yardsticks for comparison
  - Psychological lower bounds



# **Digital Library System**

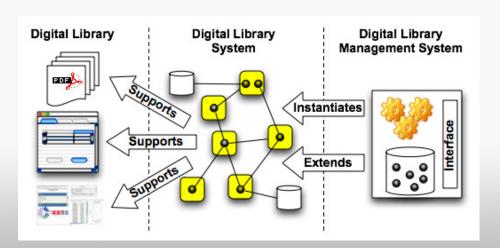
A software system that is based on a (potentially distributed) architecture and provides all functionality that is required by a particular Digital Library. Users interact with a Digital Library through the corresponding Digital Library System.





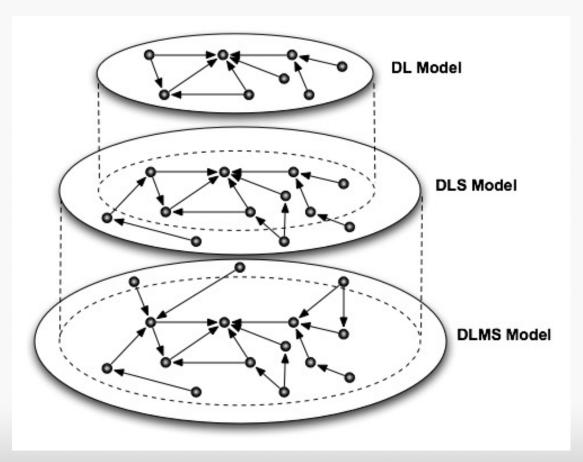
# **DL Management System**

A generic software system that provides the appropriate software infrastructure to both (i) produce and administer a Digital Library System that incorporates all functionality that is considered foundational for Digital Libraries and (ii) integrate additional software offering more refined, specialized, or advanced functionality.





# Hierarchy of Conceptualizations





## **Main Roles of Actors**

End-Users



DL Designers



DL System Administrators



DL Application Developers





## **Main Roles of Actors**

DL End-Users



DL Administrators



DL Software Developers





## **DL End Users**

Exploit the DL functionality for providing, consuming, and managing DL Content and some of its other constituents.

Further partitioned into

- Content Creator
- Content Consumer
- Digital Librarian



# **Digital Librarian**

Select DL content and its sources

Catalogue DL content

Curate DL content

Select supporting content, e.g., repositories, ontologies, authority files, ...

Establish all resources, e.g., policies, quality, ...



## **DL** Designers

Exploit their knowledge of the application semantic domain to define, customize, and maintain the Digital Library so that it is aligned with the information and functional needs of its endusers. To perform this task, they interact with the DLMS providing functional and content configuration parameters.





## **DL System Administrators**

Select the software components necessary to create the Digital Library System needed to serve the required DL and decide where and how to deploy them. They interact with the DLMS by providing architectural configuration parameters, such as the selected software components, the hosting nodes, and the components allocation.





## **DL Administrators**

## Interact with DLMS

Define, customize, and maintain DL
Provide functional & content configuration parameters
Content format
User profile format
Document model

Select software components

Decide where and how to deploy them

Provide architectural configuration parameters

Match software components with hosting nodes

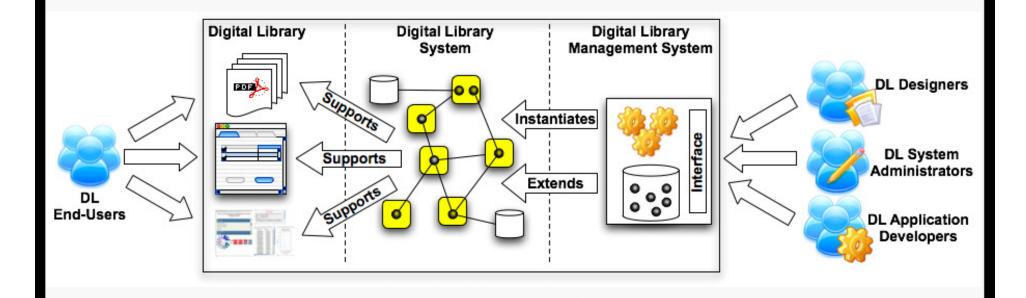


# **DL Software Developers**

Develop software components for DLMSs and DLSs



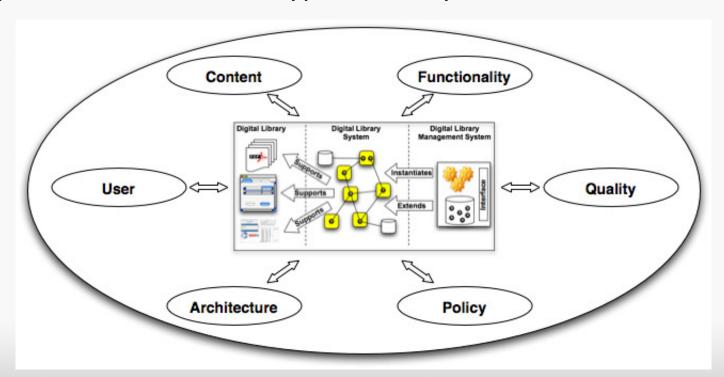
## The user's views





## The Model

Concepts and relationships that represent the significant aspects of the different type of DL "systems"





# The Digital Libraries Reference Model in a Nutshell



## 3 types of systems

- •DL
- •DL System
- •DL Management Systems

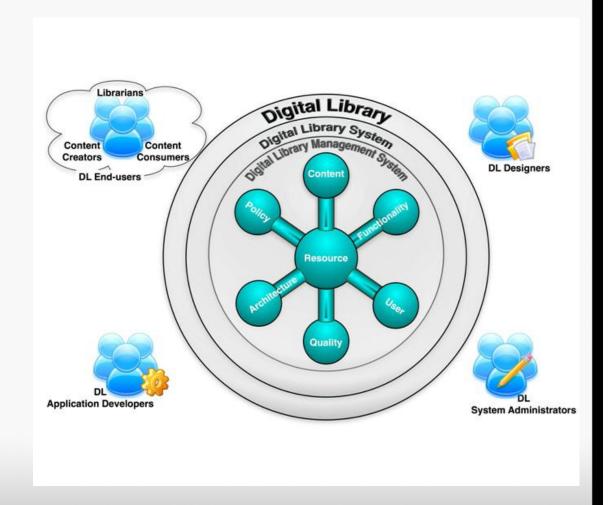
## • 6 (7) +1 Domains

- (Organization)
- Content
- User
- Functionality
- Policy
- Quality
- Architecture
- + Resource

### 4 Role of Actors

- DL end-Users
- DL Administrators
- DL Software Developers

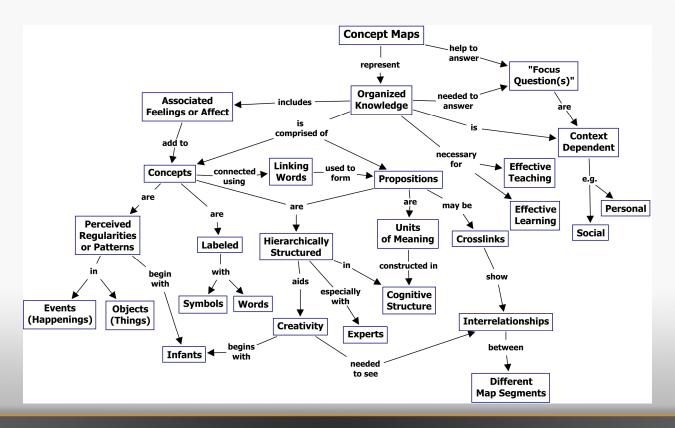
## The DL Universe





# **Concept Maps**

The Reference Model describes the Digital Library Universe **by Concept Maps** i.e. graphical tools for organizing and representing knowledge in terms of **Concepts** and **Relationships** 





## **Concepts and Relationships**

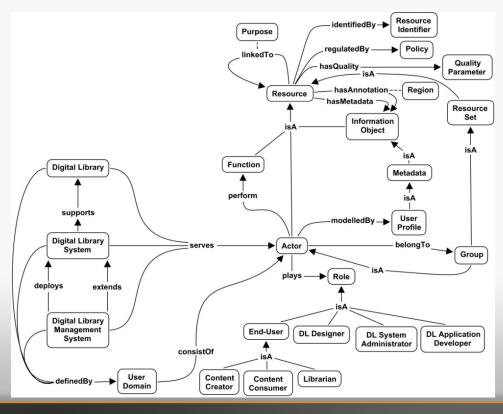
**Definition:** A set of *Actors* sharing certain characteristics, that may interact with one another, accept expectations and obligations as members of the group, and share a

common identity.

Relationships: .....

Rationale: ....

**Examples: ....** 

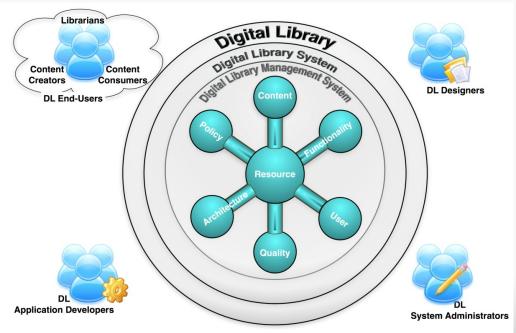




# The RM is founded on 6 + 1 Domains

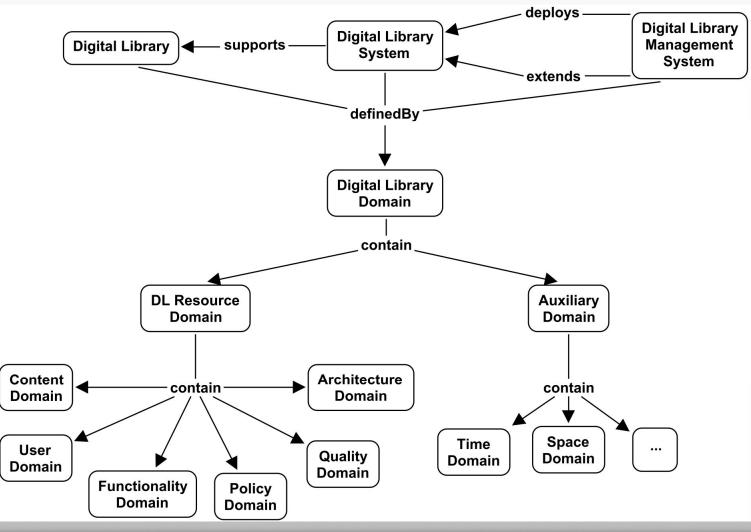
- 1. Content information available
- **2. User** actors interacting with system
- 3. Functionality operations supported
- **4. Policy** rules and conditions governing operation
- **5. Quality** qualitative & quantitative characterisations of system
- **6. Architecture** –physical software (&hardware) constituents concretely realising the DL

**Resource** – captures generic characteristics (super-domain)



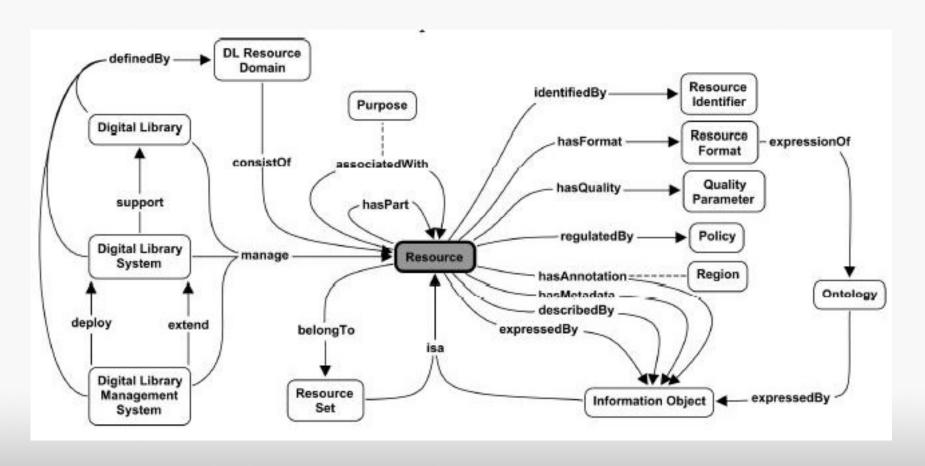


## The DL Domains





# The DL Resource Domain (1/3)





## The DL Resource Domain (2/3)

#### **Resource Domain**

- At the highest-level
- Represents all entities and relationships managed in DL Universe

#### Resource

- Most general concept of the DL Resource Domain
- Captures any Digital Library entity
- Can be grouped into Resource Sets



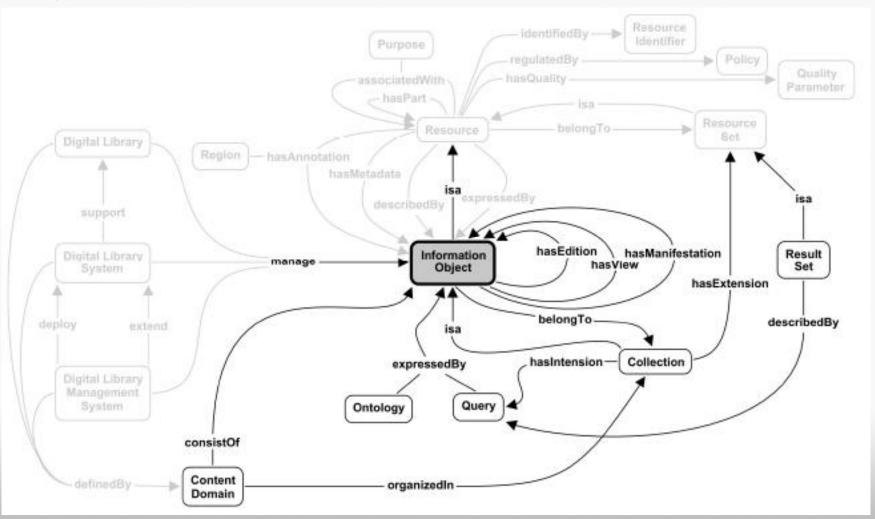
## The DL Resource Domain (3/3)

#### Each Resource is

- identified by a Resource Identifier (<identifiedBy>)
- expressed by an Information Object (<expressedBy>)
- described by or commented on by Information Objects, such as Metadata (<hasMetadata>) and Annotations (<hasAnnotation>)
- arranged or set out according to a *Resource Format* (<hasFormat>), which may be drawn from an Ontology
- characterised by Quality Parameters, each capturing how resource performs with respect to some attribute (<hasQuality>)
- regulated by **Policies** (<*regulatedBy*>) governing all aspects of its lifetime



## **The Content Domain**





## The Content Domain (cont.)

#### **Content Domain**

represents all entities related to the DL information

Information Object is a Resource Information Object includes

- Text documents
- Images, sound, multimedia, 3-D objects, games and virtual reality documents
- Data-sets, databases
- Composite objects and collections



## **The Content Domain:**

**Further classification of Information Objects (1)** 

#### By type of knowledge, information, or data

- Raw data captured directly from outside world (especially data or data streams captured by instruments)
- Data processed through or generated by the mind or some other system - often called knowledge or information (and not raw data)

#### By type of information representation or encoding

- Encoded in natural language and embodied in a document, including pictorial or sound representations
- Encoded in a formal structure, including database tables or formal entity-relationship statements and ontologies



## **The Content Domain:**

#### **Further classification of Information Objects (2)**

#### By state of digital representation

- Information object born digital (text or image)
- Information object produced by digitization of a non-digital information object
- Non-digital information object represented by metadata record

#### By level of abstraction

To choose one existing method, e.g., IFLA FRBR:

**Work**, e.g., general idea of a story , **Expression**, e.g., telling of a story in a text, **Manifestation**, e.g., graphic image showing letters and words that make up text, **Item**, e.g., an individual, physical object (e.g., printed copy) of a manifestation



# The Content Domain: example

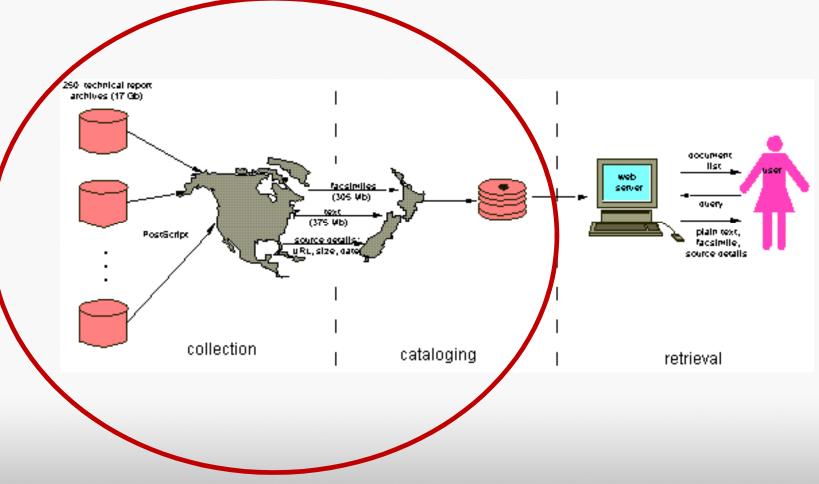
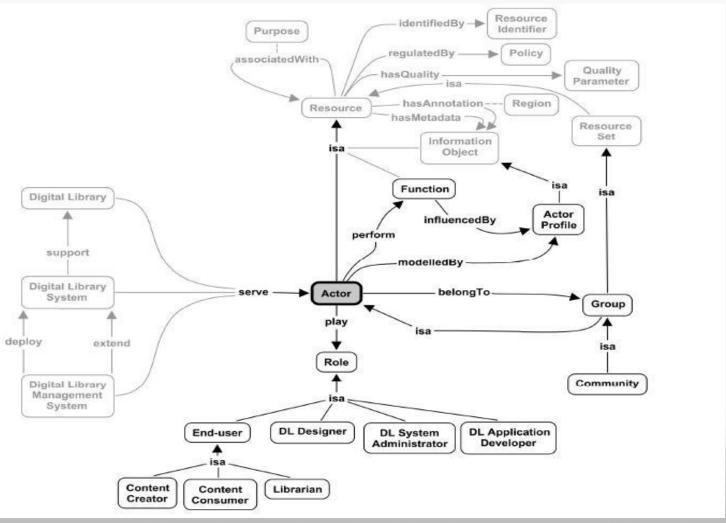


Image: Digital Libraries Based on Full-Text Retrieval, Ian H. Witten, Craig G. Nevill-Manning and Sally Jo Cunningham



## The User Domain





## The User Domain (cont.)

#### **User Domain**

represents all external entities interacting with DL humans, as well as software programs or physical instruments

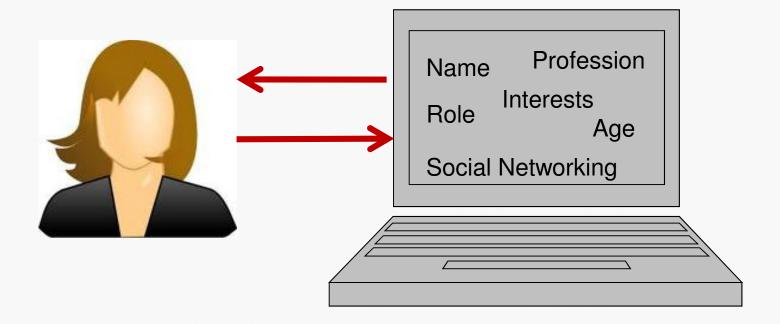
Actor is a Resource

Actor has an Actor Profile and one or more Actor Roles
Roles includes

- End User Content Creator, Content Consumer, Librarian
- DL Designer
- DL System Administrator
- DL Application Developer

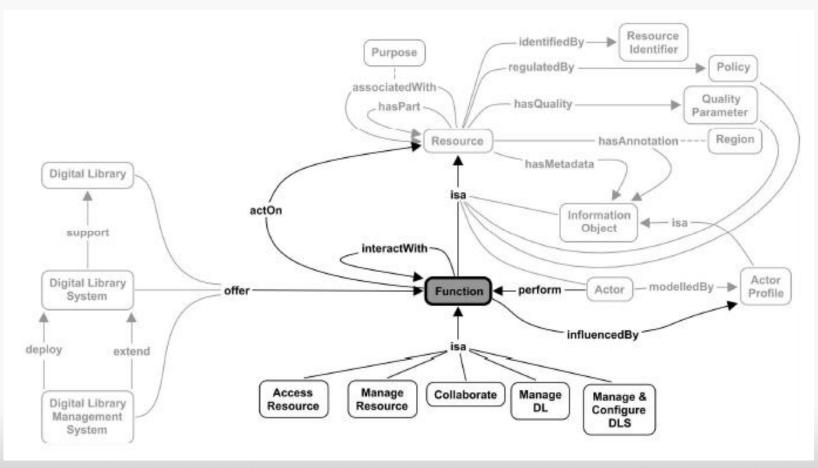


## The User Domain: example





## The Functionality Domain





Functionality Domain captures all processing on *Resources* (most often on *Information Objects*) and other necessary activitiesFunction is the most central concept

Function is a Resource

**Actors** perform **Functions** 

Main **Function** specializations

- Access Resource
- Manage Resource
- Collaborate
- Manage DL
- Manage & Configure DLS



# The Functionality Domain: example

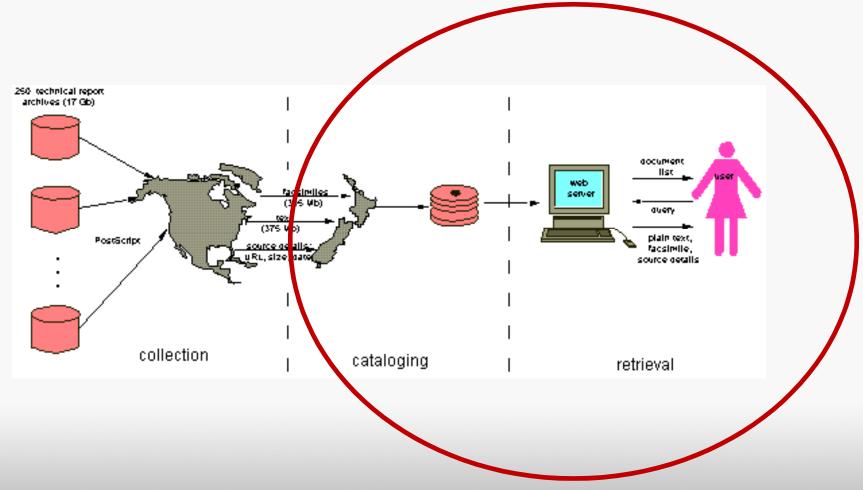
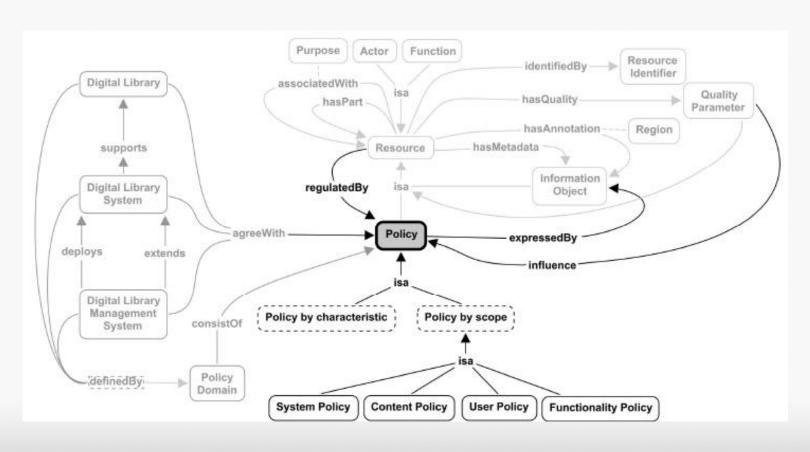


Image: Digital Libraries Based on Full-Text Retrieval, Ian H. Witten, Craig G. Nevill-Manning and Sally Jo Cunningham



## **The Policy Domain**





## The Policy Domain (cont.)

Represents the set of most critical conditions, rules, terms or regulations governing the operation of DL broad and dynamic by nature

Policy – the most central concept in the Domain Policy is a Resource

Policy by Characteristic:

Extrinsic Policy vs. Intrinsic Policy Implicit Policy vs. Explicit Policy Prescriptive Policy vs. Descriptive Policy

Policy by Type

- System Policy
- Content Policy
- User Policy
- Functionality Policy

elto3

Αυτό ίσως μπερδέψει. Είτε να το περιγράψουμε ενώ δείχνουμε το concept map για να φανεί ότι είναι διαφορετικό από το όνομα domain, είτε να το βγάλουμε;

Αξίζει να προτείνουμε στο πλαίσιο των αλλαγών που εισάγουμε στο RM, να αντικατασταθεί η λέξη αυτή από κάποια άλλη με το ίδιο νόημα; elto, 12/9/2010

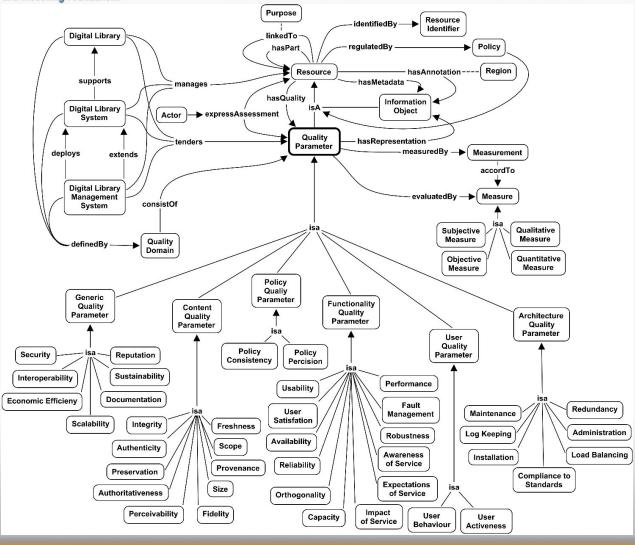


# The Policy Domain: examples

- Risk management
- Collection Development
- Object delivery
- Support
- Purchasing
- Appraisal
- User management



## **The Quality Domain**





## The Quality Domain (cont.)

Represents the aspects that need to be consider from a quality point of view in the DL

Quality Parameter: Most central concept in the Domain

Quality Parameter is a Resource

express the assessment of an **Actor**, about a **Resource** can be evaluated according to different **Measures** are actually measured by a **Measurement** 

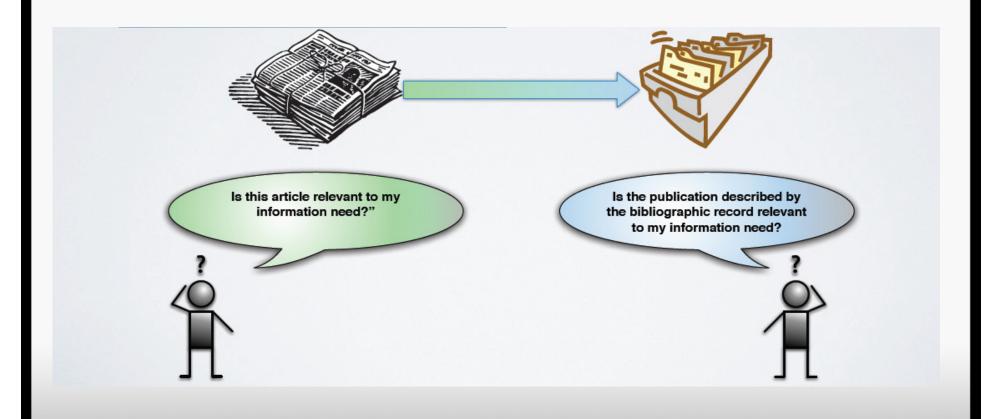
**Generic Quality Parameters** 

**System Quality Parameters** 

**Domain related Quality Parameters** (content, user, functionality, policy, architecture)

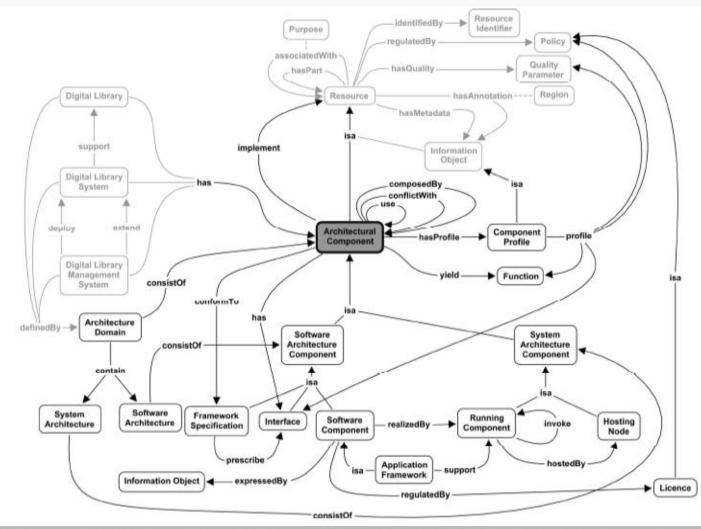


# The Quality Domain: example





## The Architecture Domain





## The Architecture Domain (cont.)

#### **Architecture Domain:**

Captures concepts and relationships characterising the two software systems playing an active role in the DL universe, i.e.

**DLSs and DLMSs** 

### **Architectural Component:**

the most central concept in the Domain

Architectural Component is a Resource

an encapsulated part of a system

Ideally a non-trivial, nearly independent, and replaceable part of a system that fulfils a clear function in the context of a well-defined architecture



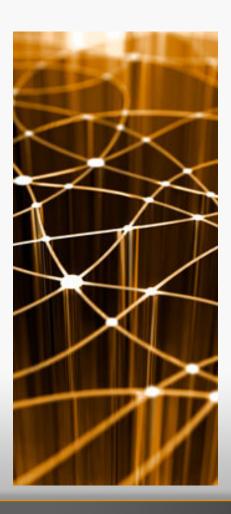
## The Architecture Domain (cont.)

#### "Component-based approach":

System assembled from discrete executable components

System may be upgraded with smaller increments, i.e., upgrading only some of constituent components
Components may be shared by systems
Though not strictly related to their being component-based, such systems tend to be distributed

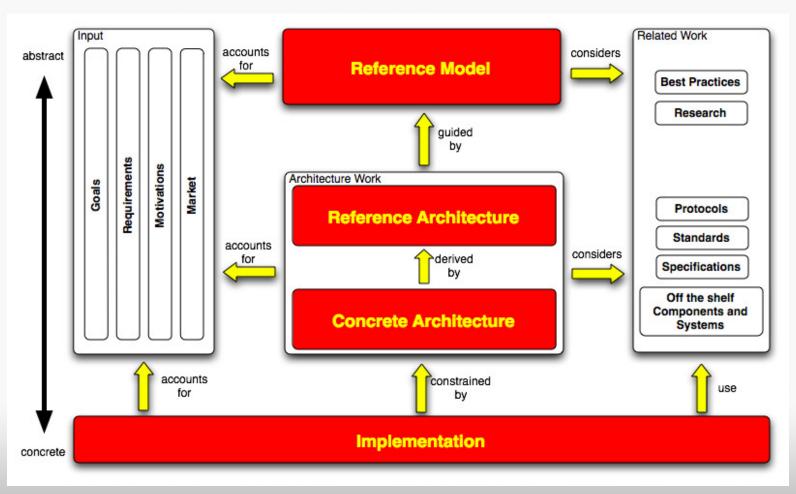




# Ευχαριστώ!

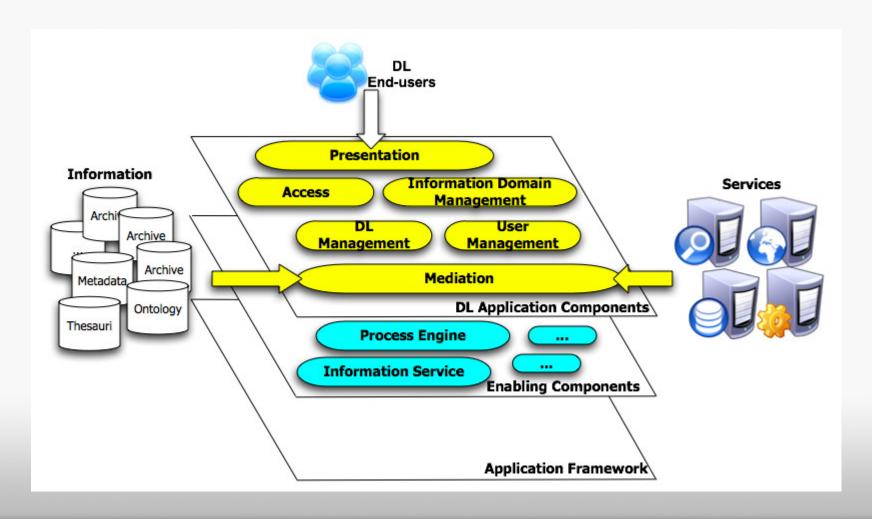


## **Reference Frameworks**



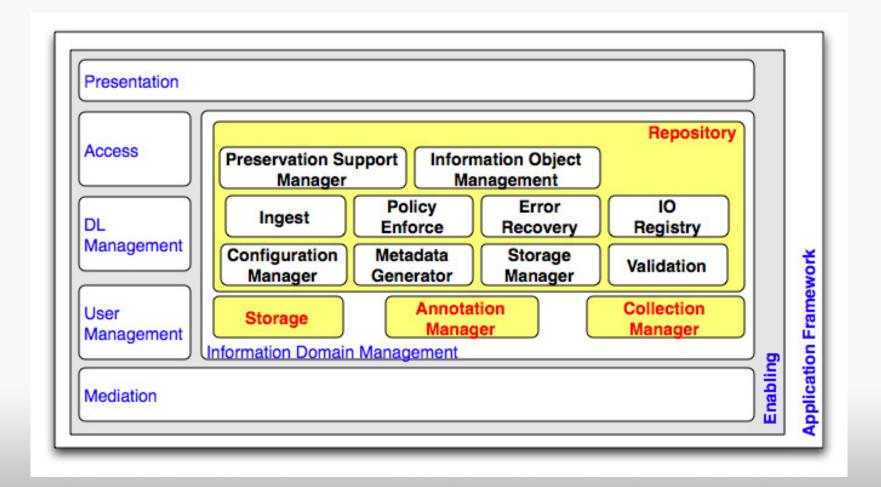


## Reference Architecture Functional Areas

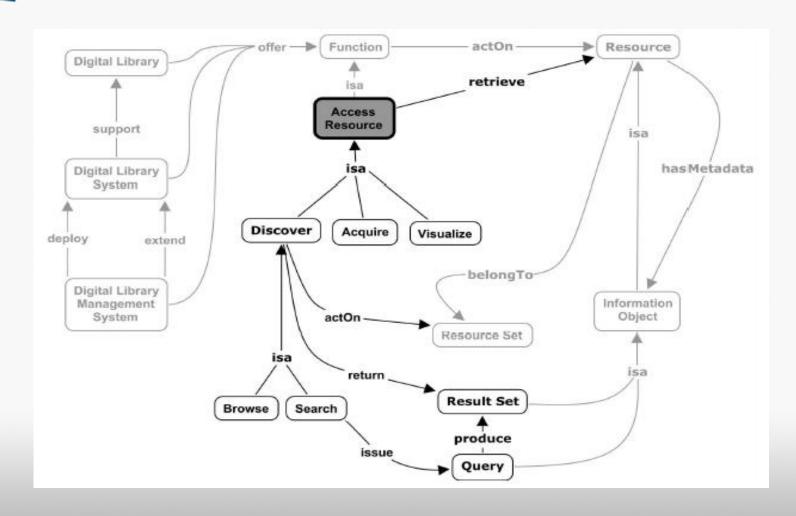




## **Functional components**



# DL.Org The Functionality Domain(3/11) Best Practices and Modelling Foundations





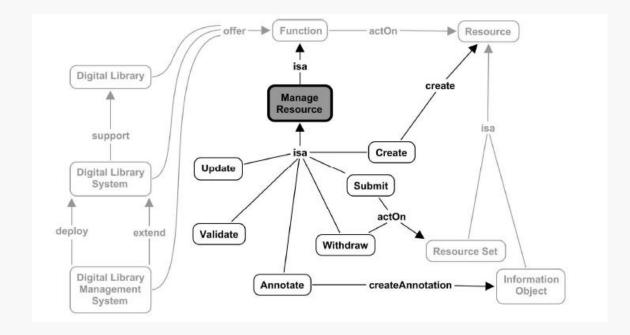
#### **Access Resource** Functions

All activities are related to requesting, locating, retrieving and finally delivering *Resources* to requestor Do not modify the DL or convert its Resources - extract some of its content and deliver it to user In most cases act on Resource Metadata



#### **Manage Resource**

General Functions
that may be applied
on all Resources
These Functions
may be specialized
for the particular
domains

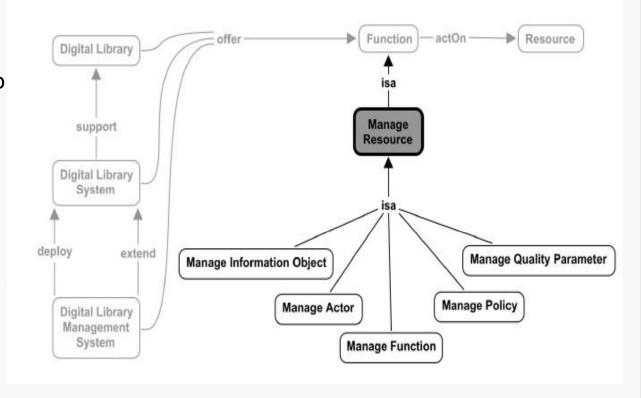


Contains general families of Functions for managing individual Resource Types



#### **Manage Resource**

- All activities related to management of Resources:
  - Creation, update, deletion
  - Analysis
  - Conversions and Transformations



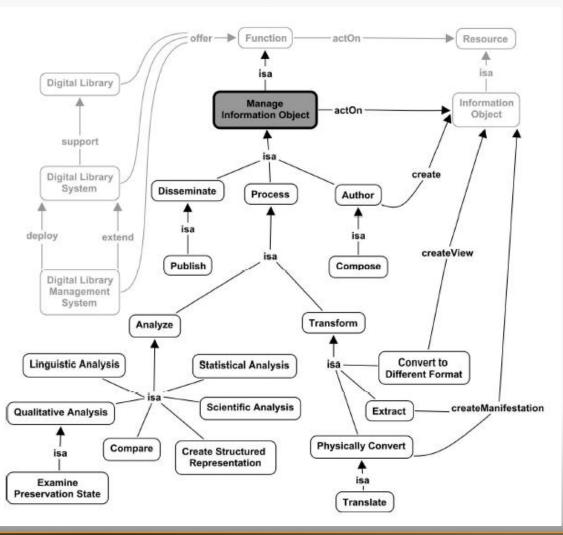


#### **Manage Information Object**

Captures creation, processing, transformation

primary Information Objects

other Information Objects or Resources in general





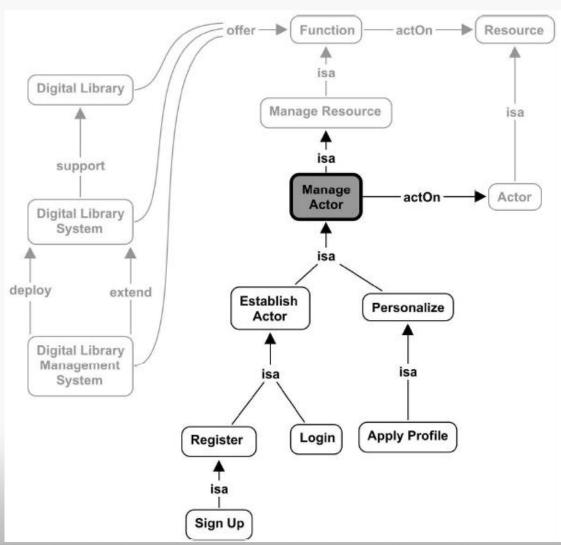
#### **Manage Actor**

Captures management of individual Actors

registration & subscription

Login

personalization of allowed functionality



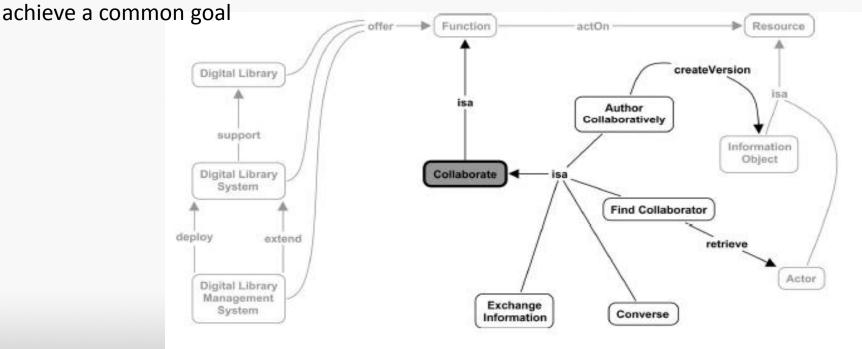


#### **Collaborate**

Captures all activities that allow multiple

Actors to work together through a DL to

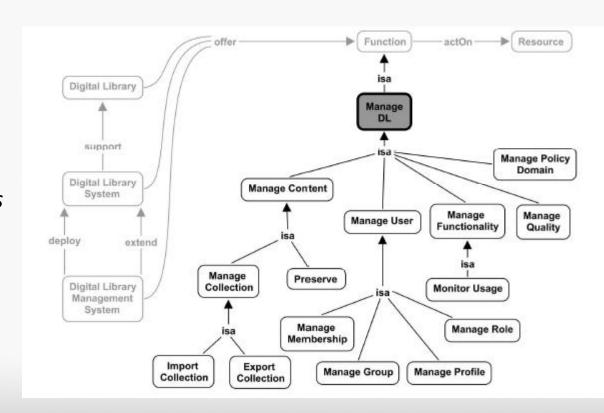
Actors to work together through a DL to





#### Manage DL

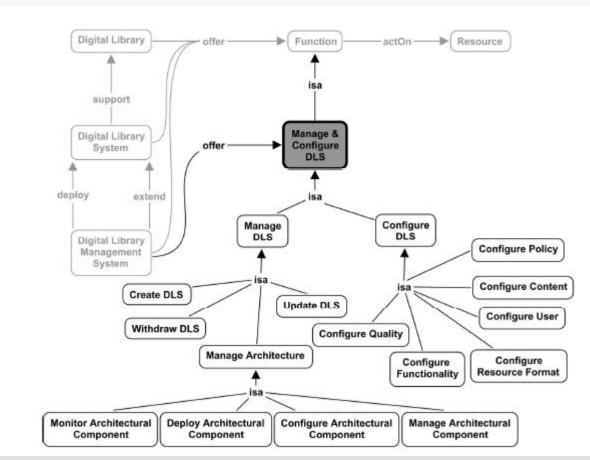
supports the day-by-day DL management, concerning all DL domains





# Manage and Configure DL

supports setting up, configuring, Digital Libr monitoring the DL





## The Quality Domain (3/4)

#### **Quality Parameter Groups:**

**Generic Quality Parameters** apply to any kind or most kinds of resources

**System Quality Parameters** apply to *Digital Library*, or a *Digital Library System*, or a *Digital Library Management System*.

Content Quality Parameters apply to Resources in the Content Domain, primarily Information Objects.

Functionality Quality Parameters apply to Resources in the Functionality Domain, primarily Functions.



## The Quality Domain (4/4)

#### **Quality Parameter Groups:**

**User Quality Parameters** apply to *Resources* in the *User Domain*, primarily *Actors* 

**Policy Quality Parameters** apply to *Resources* in the *Policy Domain*, primarily *Policies* 

Architecture Quality Parameters apply to Architectural Components, i.e., Resources belonging to the Architecture Domain



# From DELOS Reference Model to DL.org Outcomes

- Enhanced and expanded <u>Digital Library Reference</u> <u>Model, V1.0</u>, further enhancements foreseen as a part of an ongoing process
- The State of the Art Survey
- Technology and Methodology Cookbook