DL.org

Digital Library Interoperability,
Best Practices and Modelling Foundations

The DL Reference Model

Yannis Ioannidis

4 October 2010

DL.org Autumn School – Athens, 3-8 October 2010

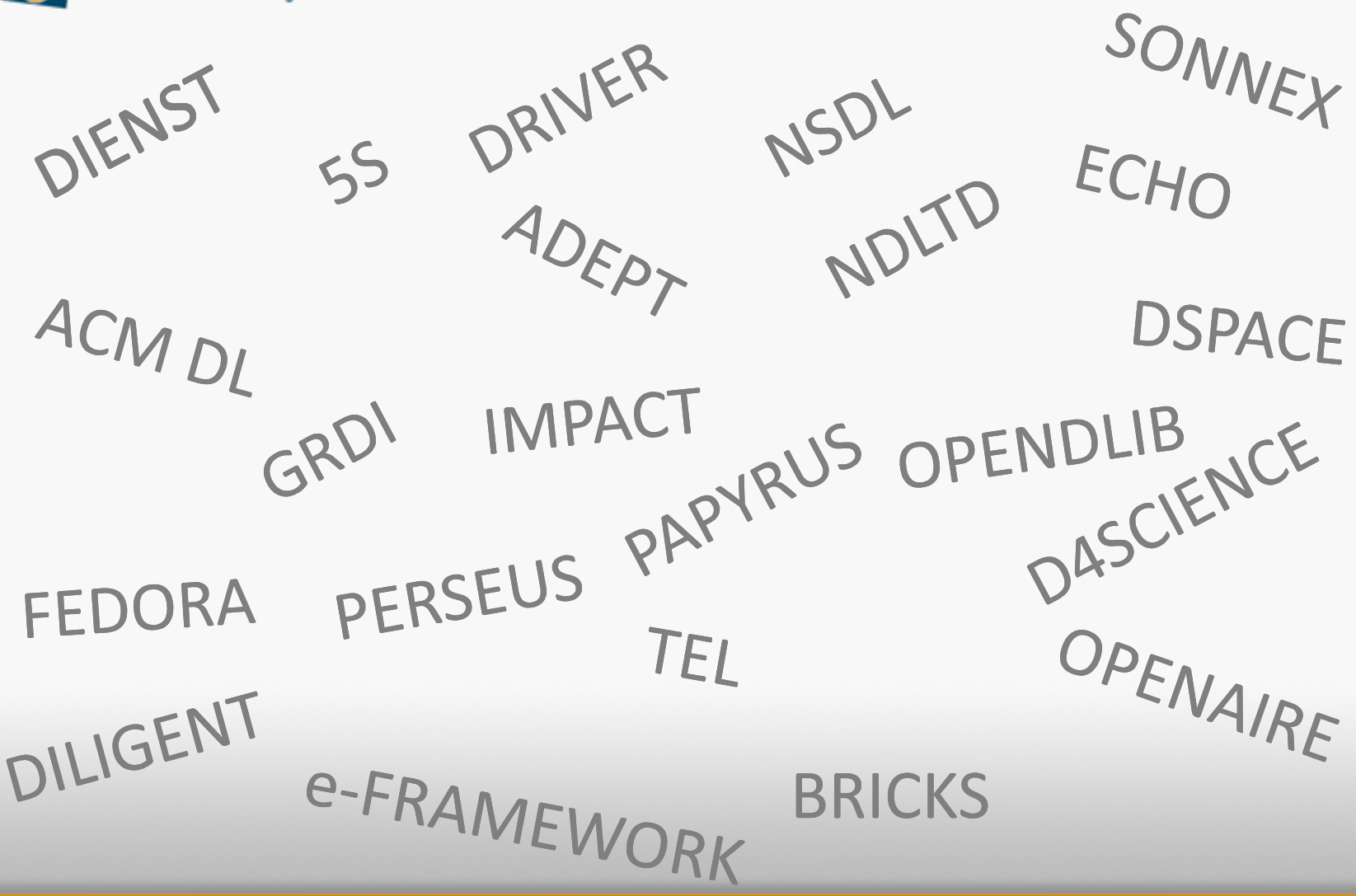
DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

Outline

- ❑ Motivations
- ❑ RM Overview
- ❑ The RM Domains
- ❑ Discussion

DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

DL Universe



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**!!!

Issues when dealing with DL

- 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

DL Reference Model

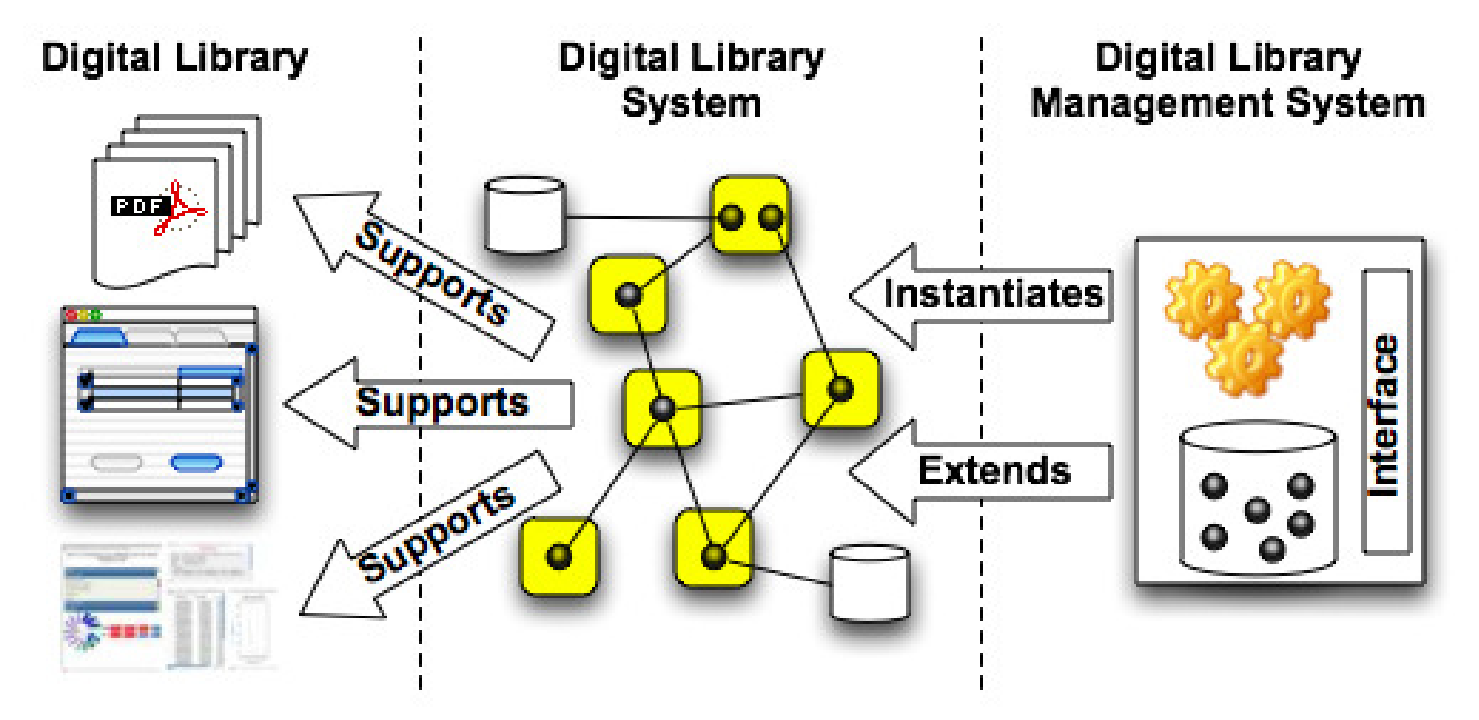
Objective

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.

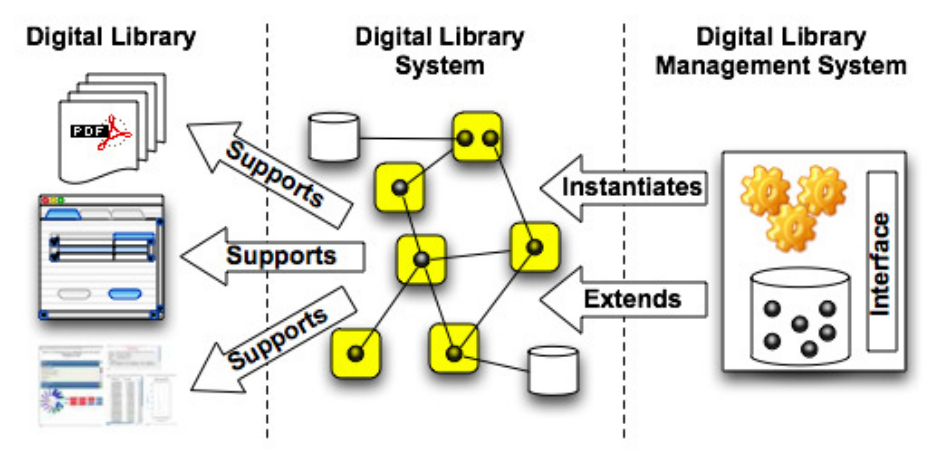
The DL "Systems"



Digital Library

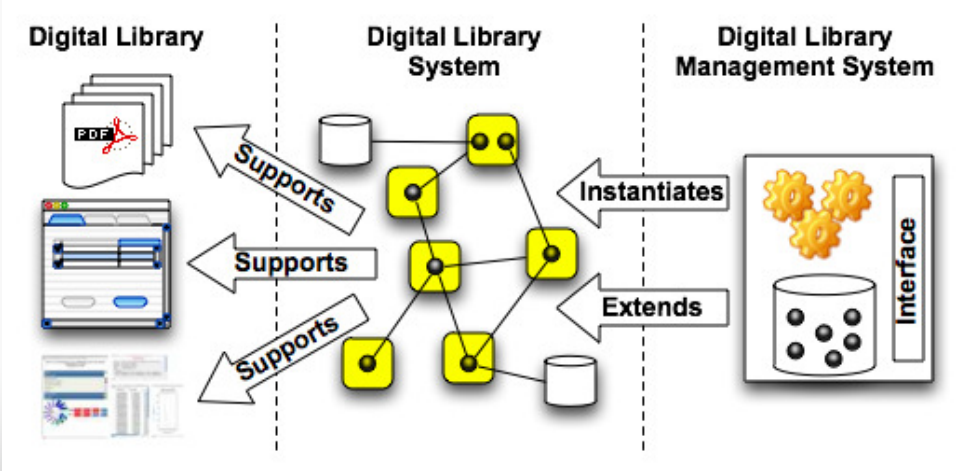
A (potentially virtual) organization that comprehensively collects, manages and preserves for the long term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies.

The *DELOS Digital Library Reference Model*



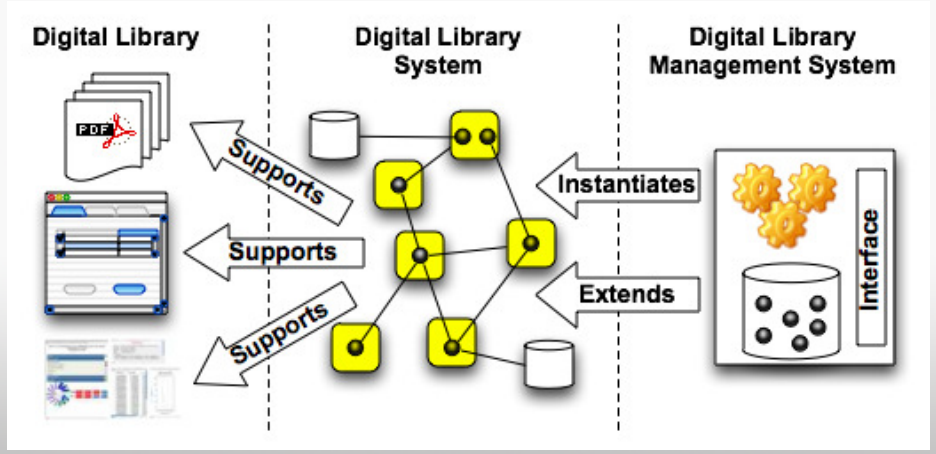
Digital Library System

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.*

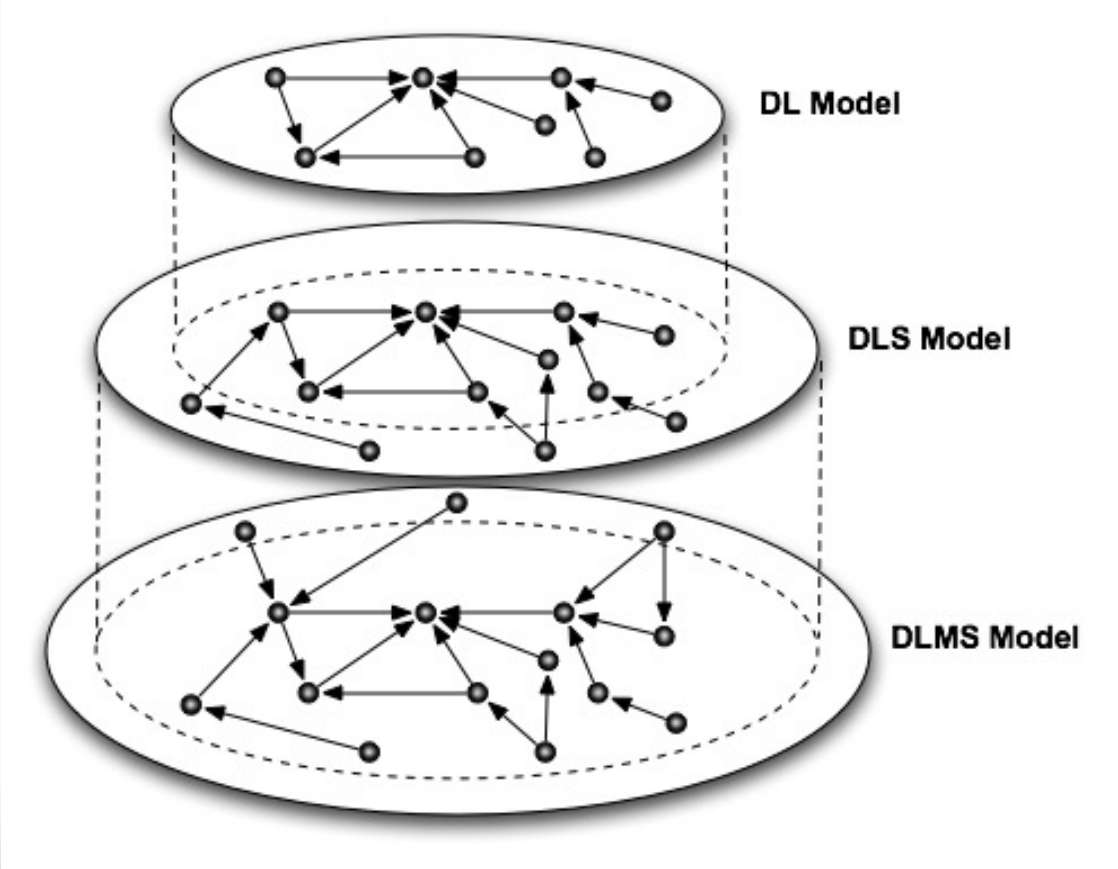


DL Management System

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

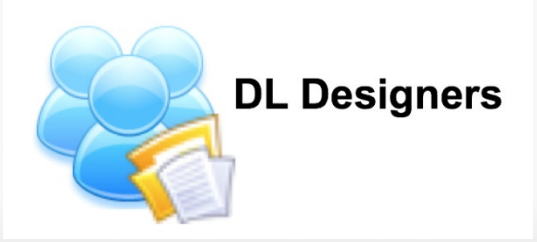
DL end users

Exploit the DL functionality for providing, consuming, and managing the DL Content as well as some of its other constituents. They perceive the DL as a statefull entity that serves their functional needs. DL end-users may be further partitioned into

- *Content Creator*
- *Content Consumer*
- *Librarian*

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 end-users. To perform this task, they interact with the DLMS providing functional and content configuration parameters.



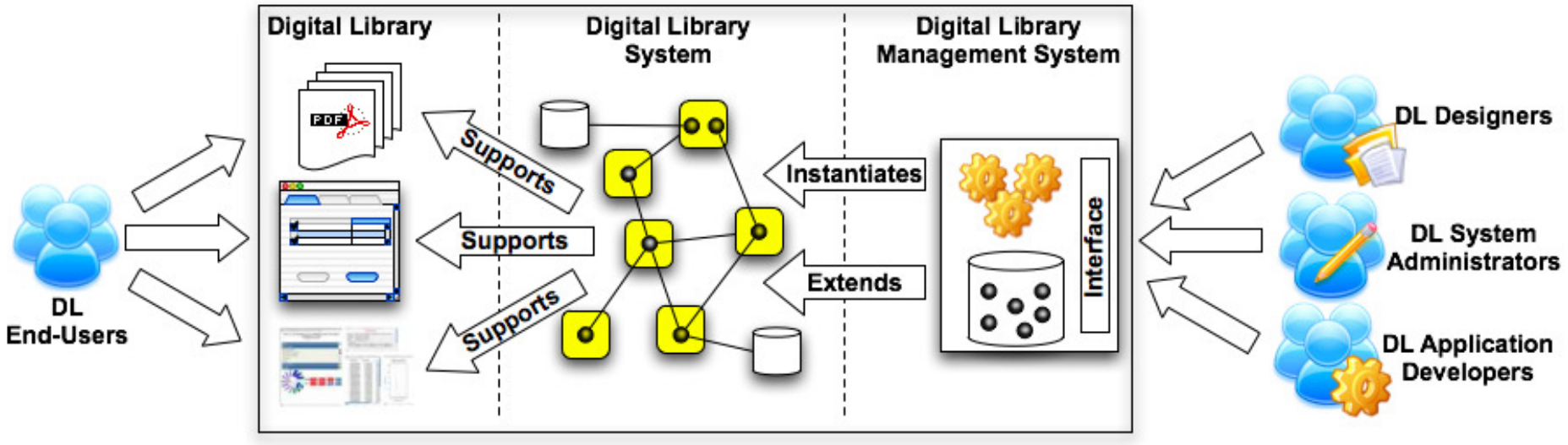
DL System Administrators

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 Application Developers

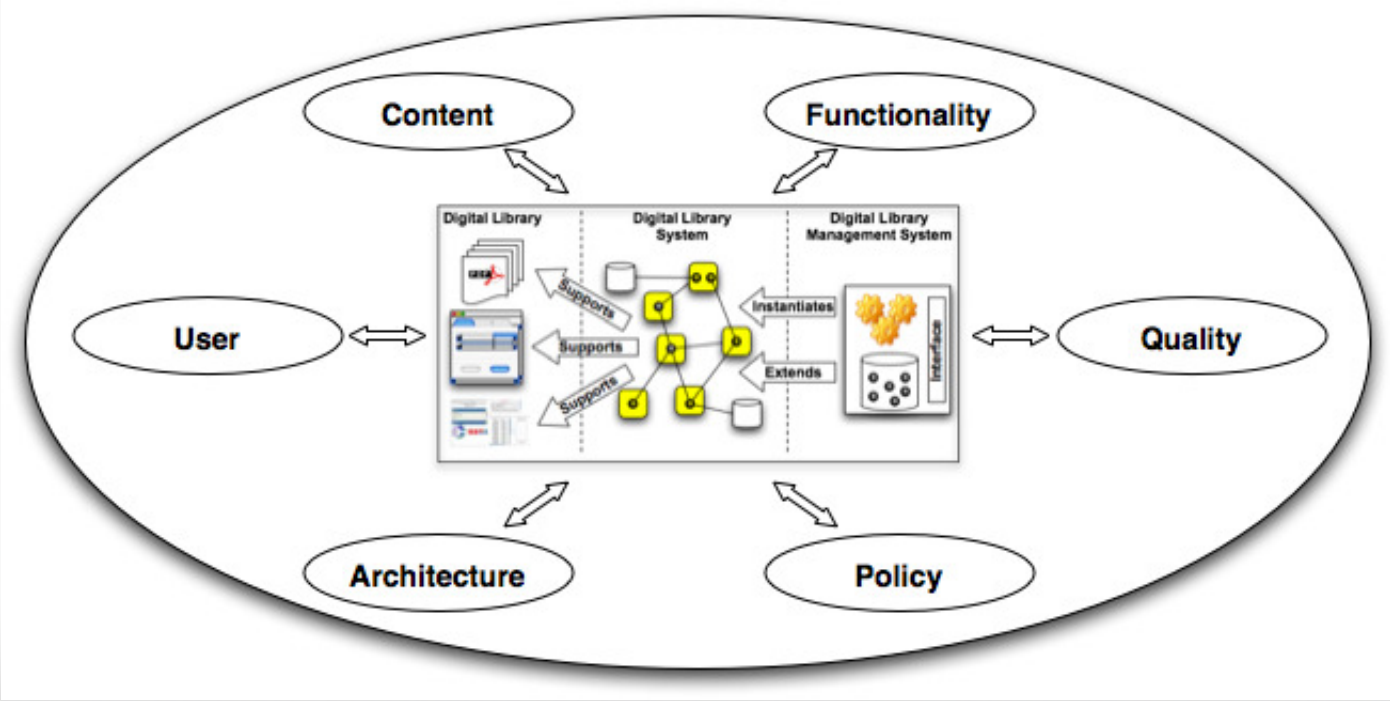
These develop the software components of DLMSs and DLSs, realizing the necessary functionality

The user's views



The Model

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



DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

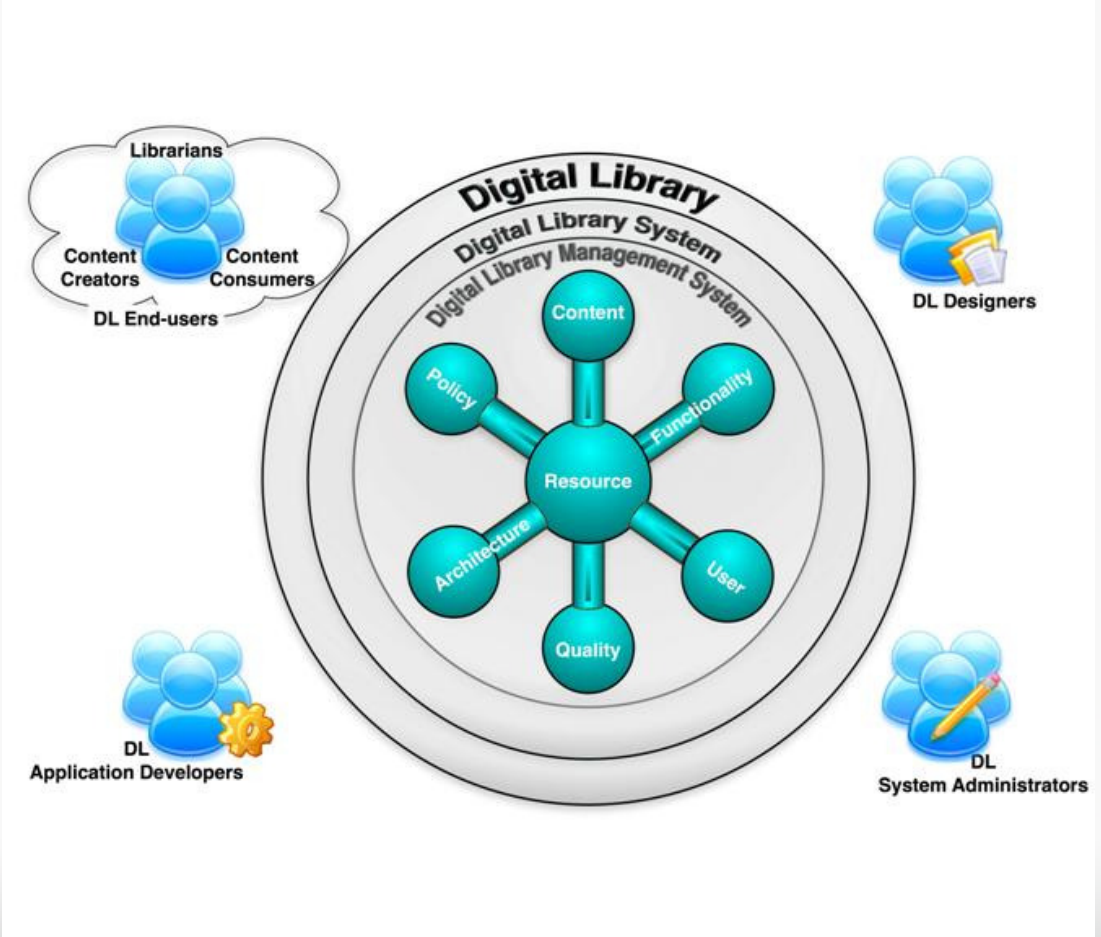
The DELOS DL Reference Model in a Nutshell

The DL Universe

3 types of systems

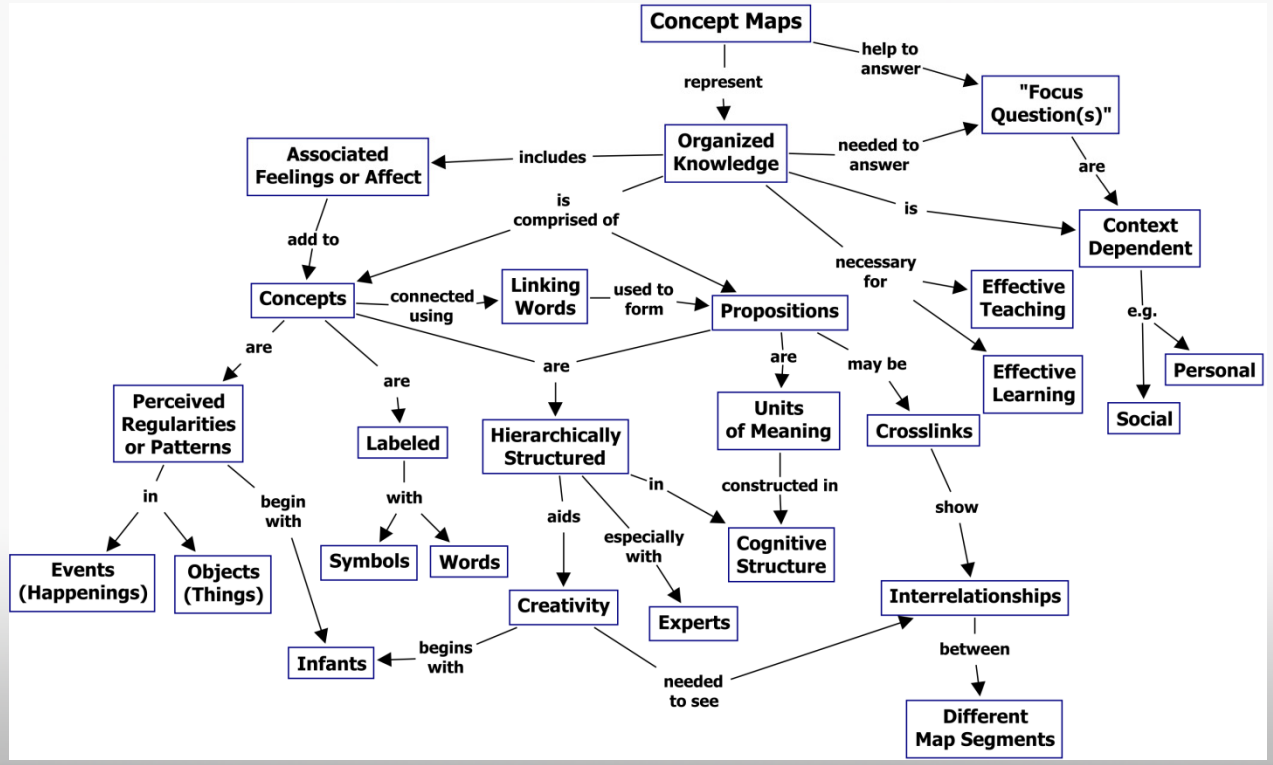
- DL
- DL System
- DL Management Systems

- **6 +1 Domains**
  - Content
  - User
  - Functionality
  - Policy
  - Quality
  - Architecture
  - + Resource
- **4 Role of Actors**
  - DL end-Users
  - DL Application Developers
  - DL Designers
  - DL Systems Administrators



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**



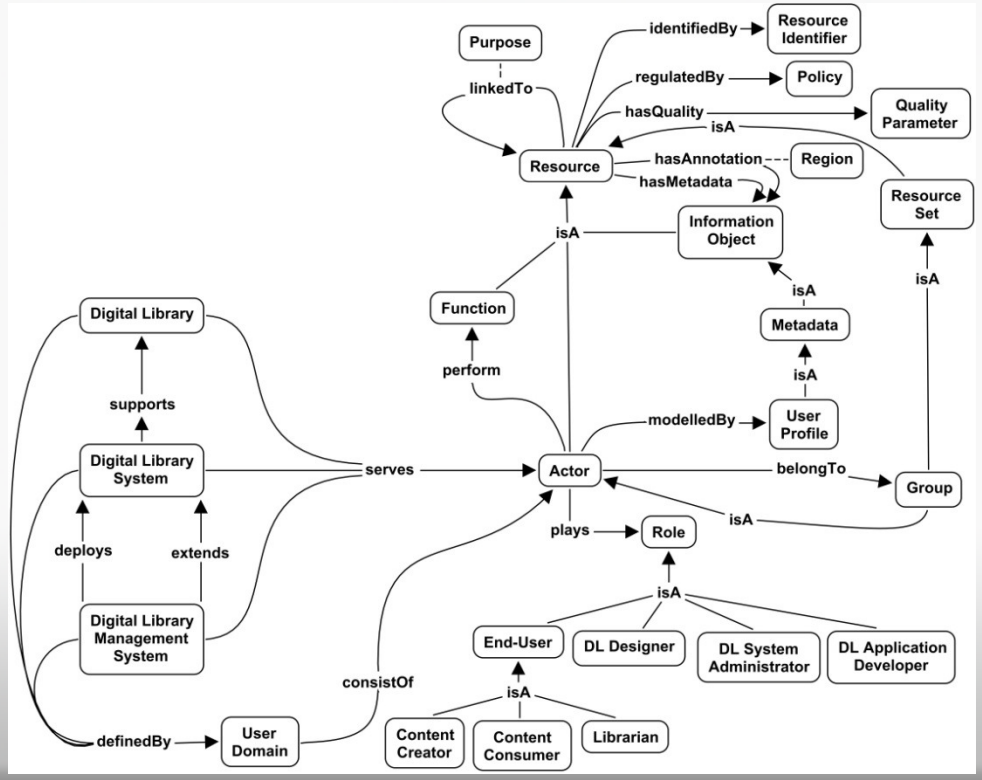
Concepts and Relationships

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.

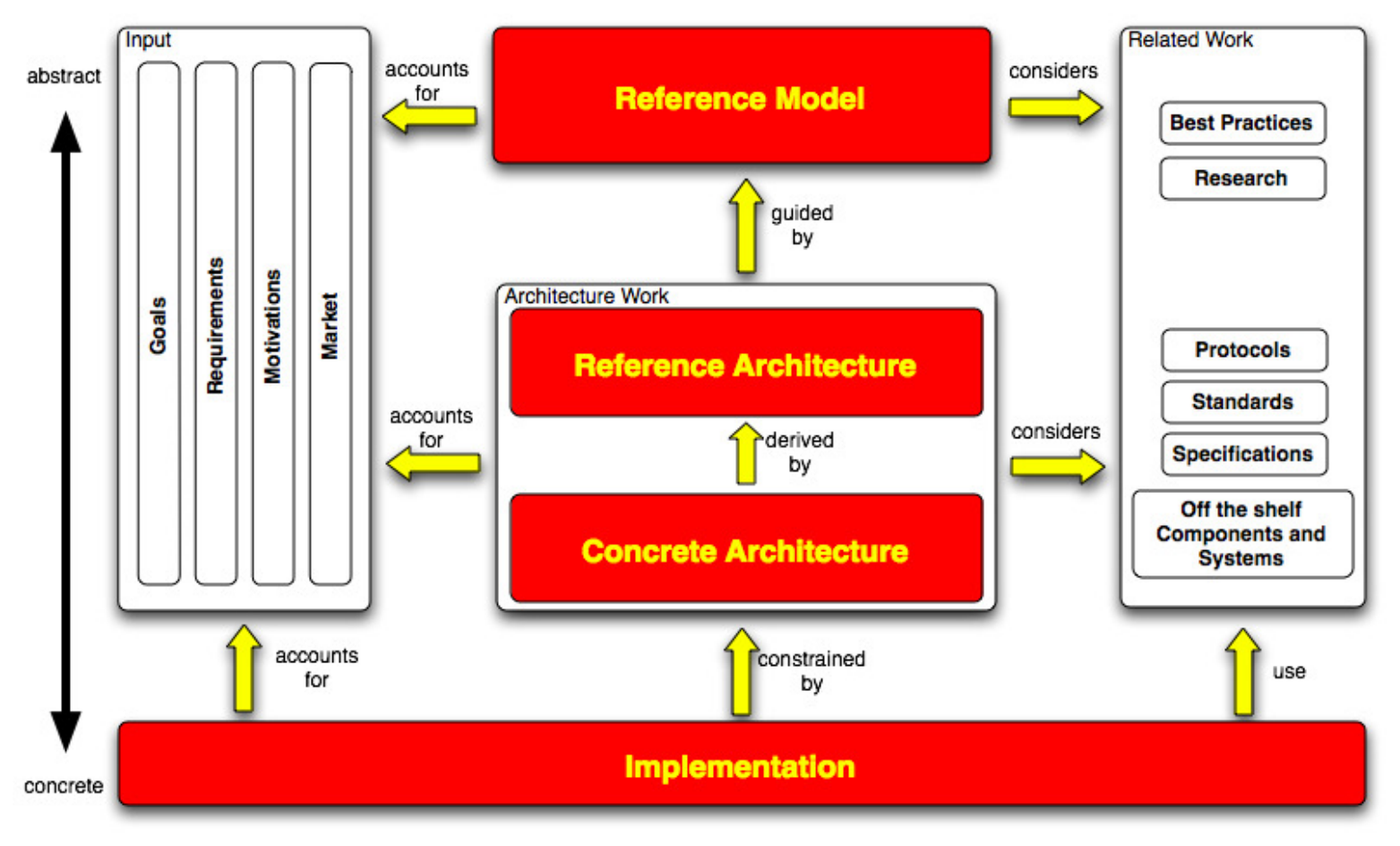
Relationships: ……

Rationale: ….

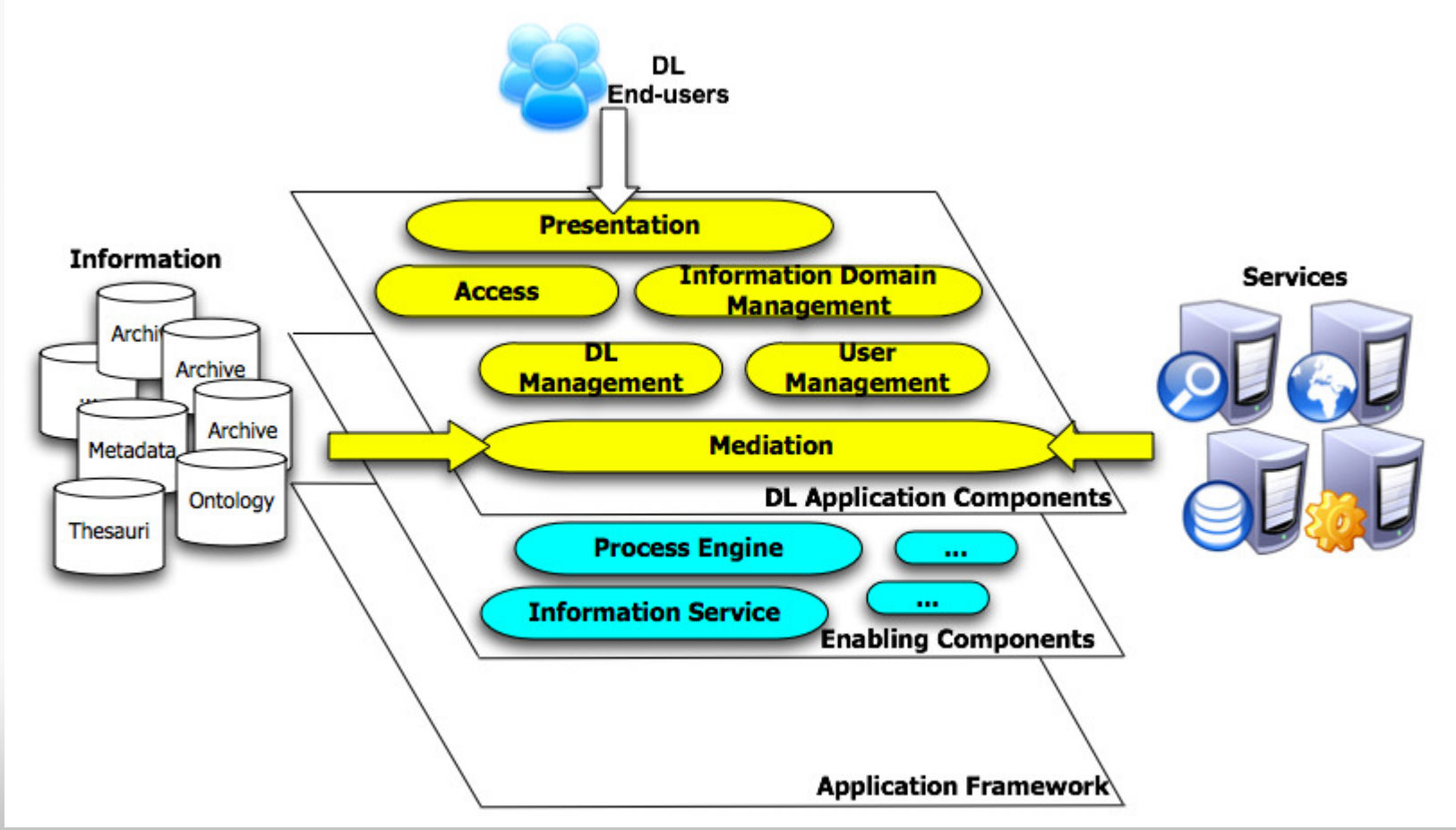
Examples: ….



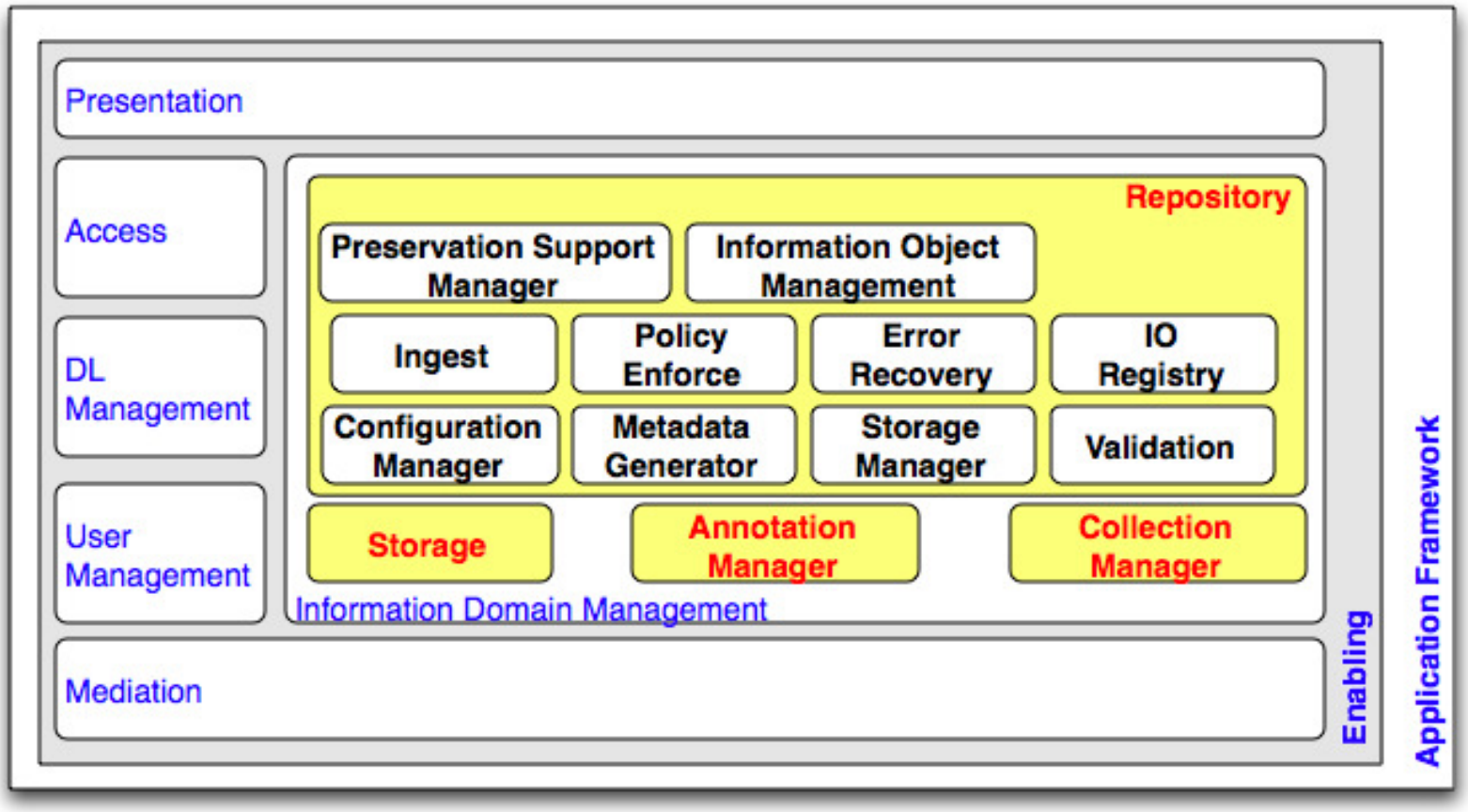
Reference Frameworks



Reference Architecture Functional Areas



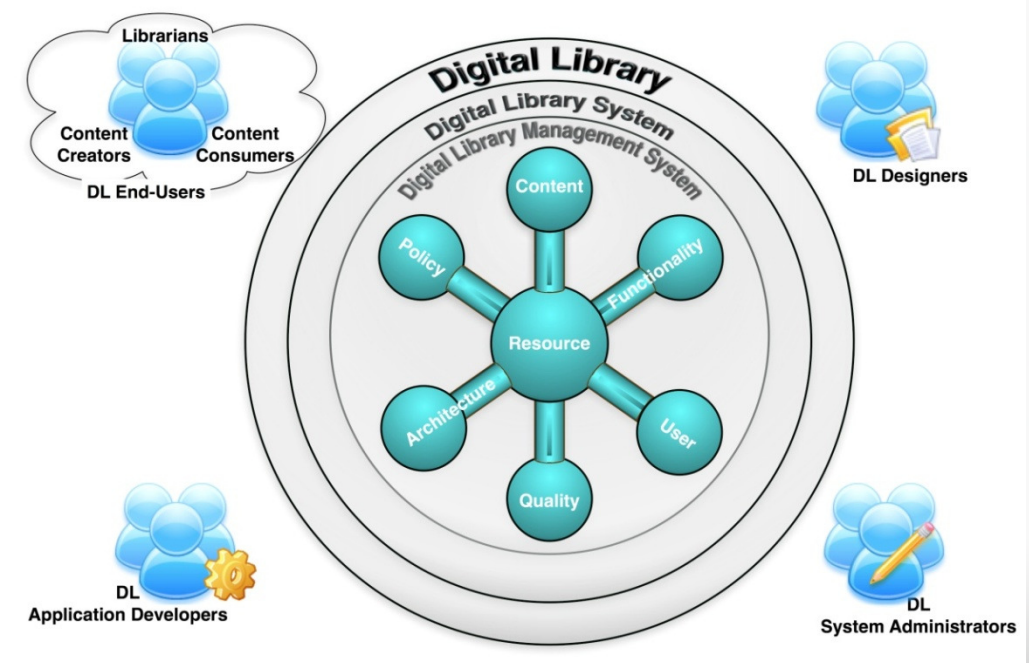
Functional components



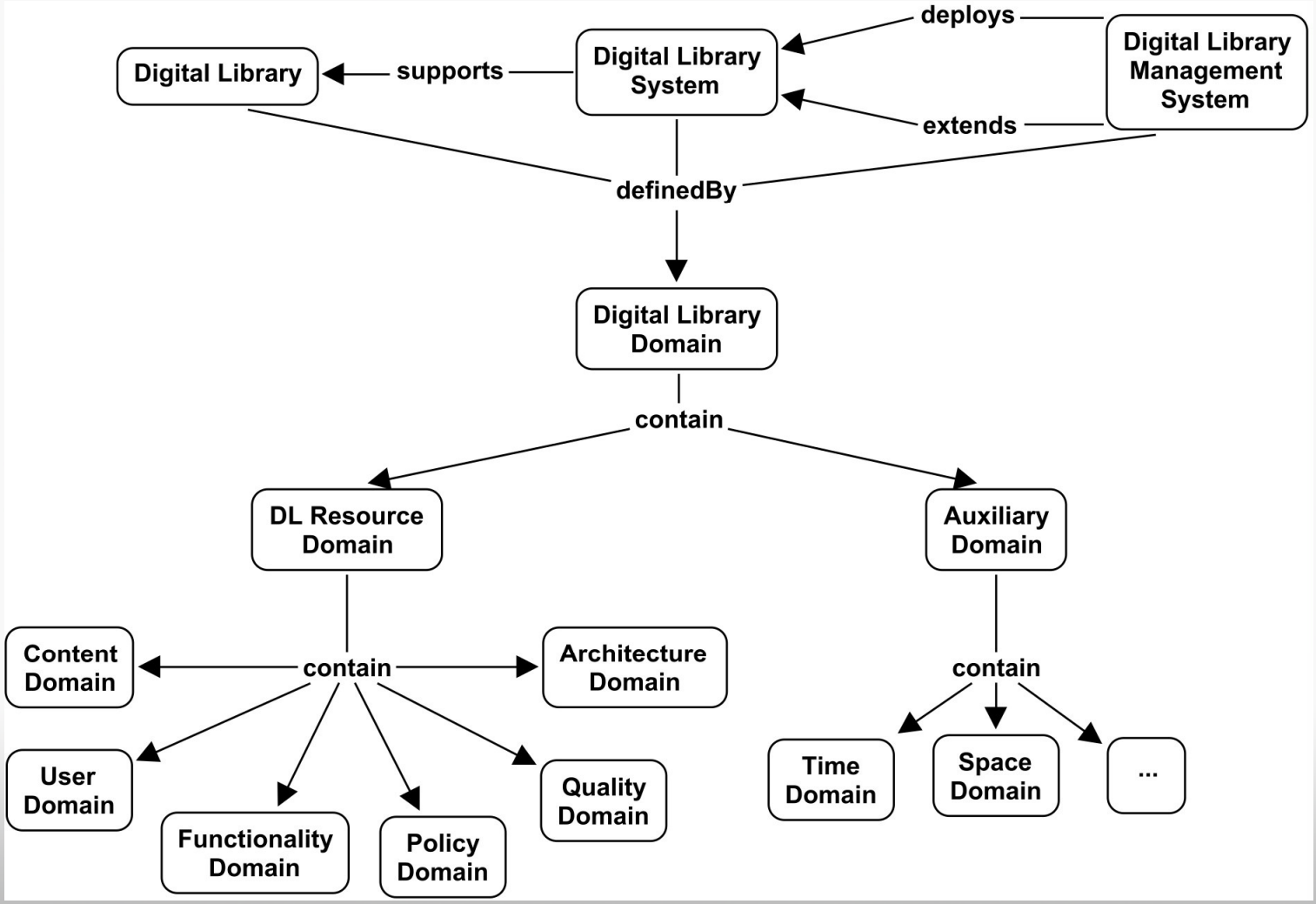
The RM is founded on 6 + 1 Domains

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

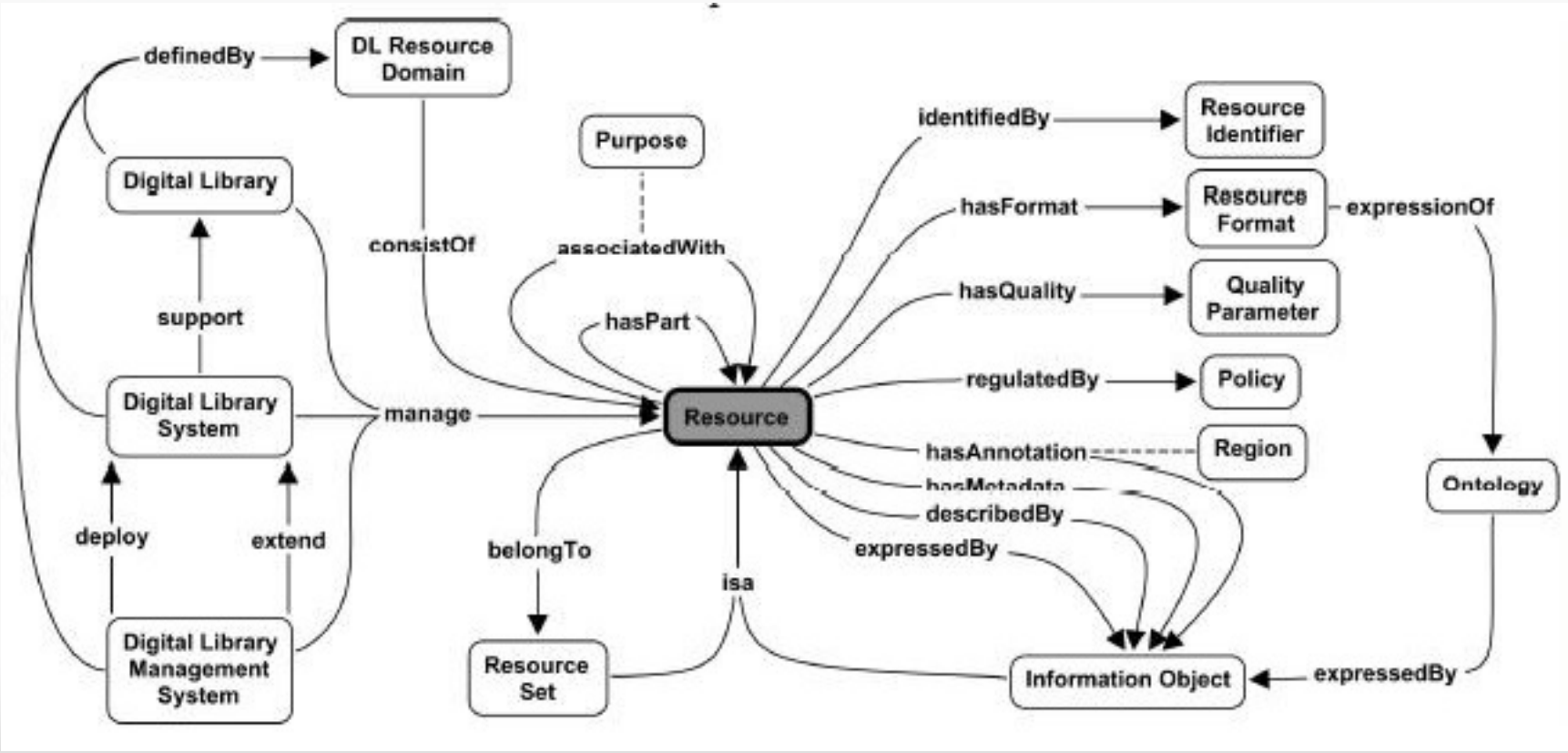
Resource – captures generic characteristics (super-domain)



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

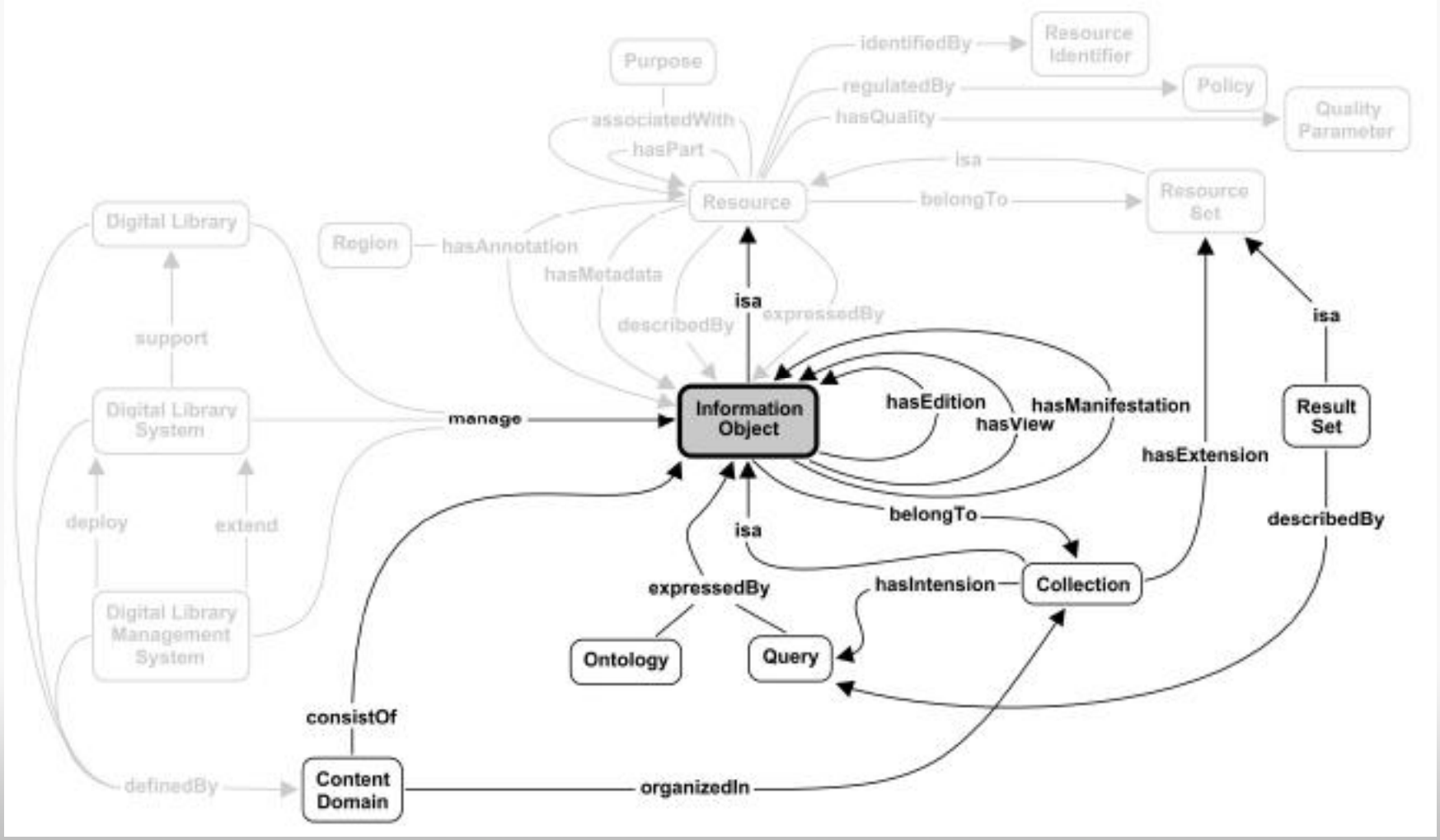
- Most general concept of the *DL Resource Domain*
- Captures any Digital Library entity
- Can be grouped into **Resource Sets**

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 (1/6)



The Content Domain (2/6)

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 (3/6)

Further classification of Information Objects

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)

The Content Domain (4/6)

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 (5/6)

By state of digital representation

- Information object born digital, such as a born digital text or a digital camera image
- Information object produced by digitization of a non-digital information object
- Non-digital information object represented by metadata record

The Content Domain (6/6)

By level of abstraction

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

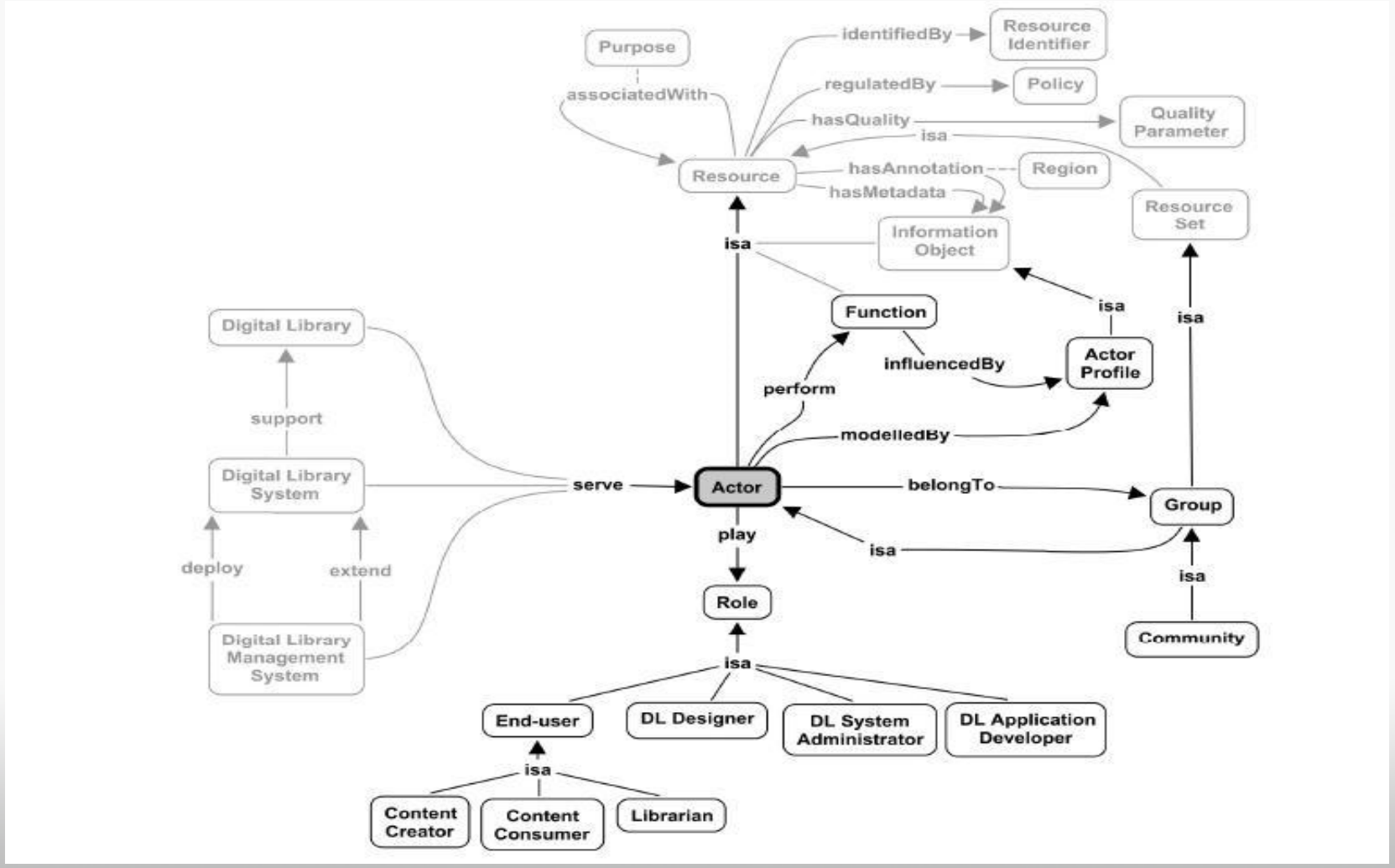
Work, e.g., general idea of a story

Expression, e.g., telling of a story in a text

Manifestation, e.g., graphic image showing letters and words that make up text - common to all copies "printed" from the same typeset image

Item, e.g., an individual, physical object (e.g., printed copy) of a manifestation

The User Domain (1/2)



The User Domain (2/2)

User Domain

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

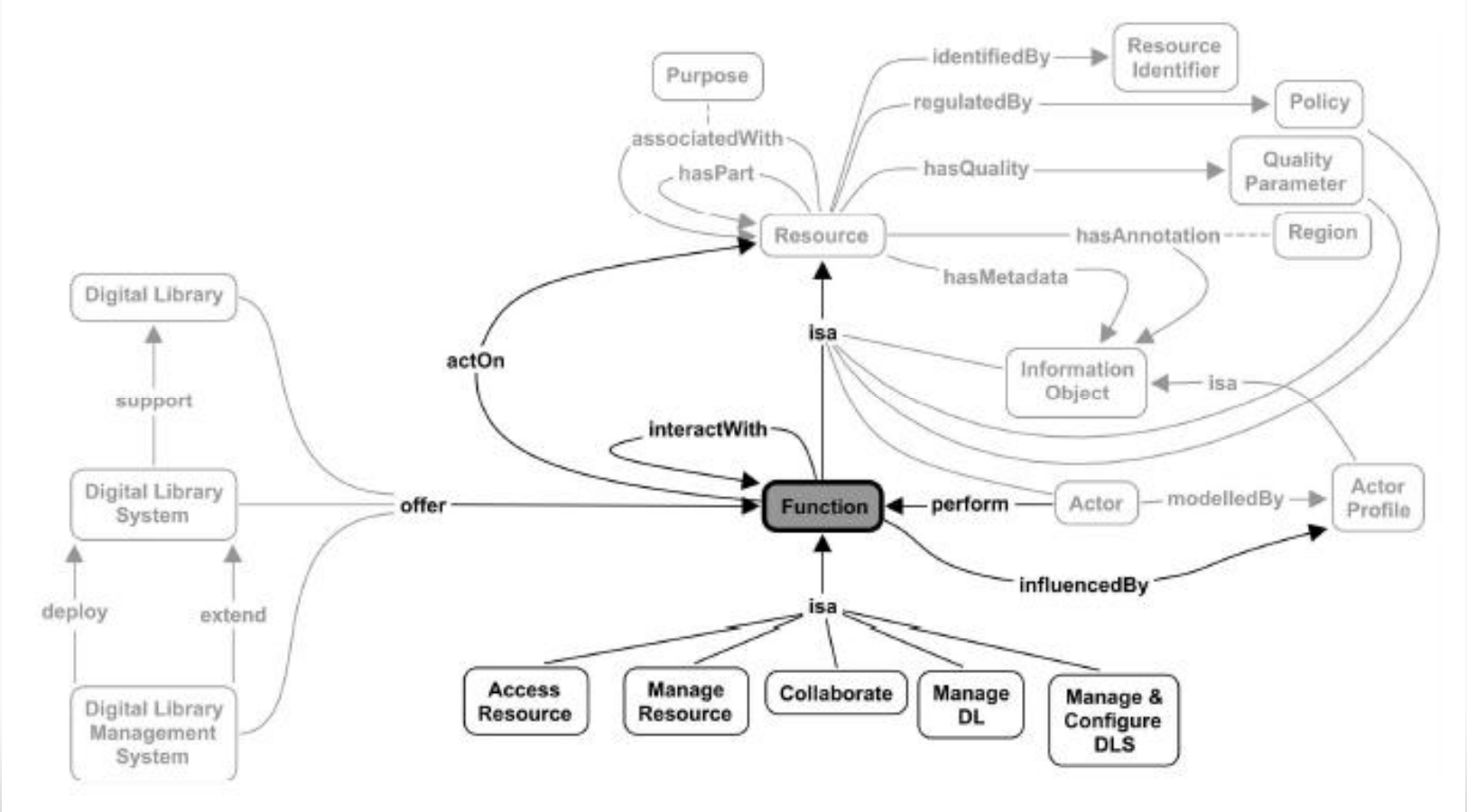
Actor is a **Resource**

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 Functionality Domain (1/11)



The Functionality Domain(2/11)

Functionality Domain captures all processing on *Resources* (most often on *Information Objects*) and other necessary activities

Function is the most central concept

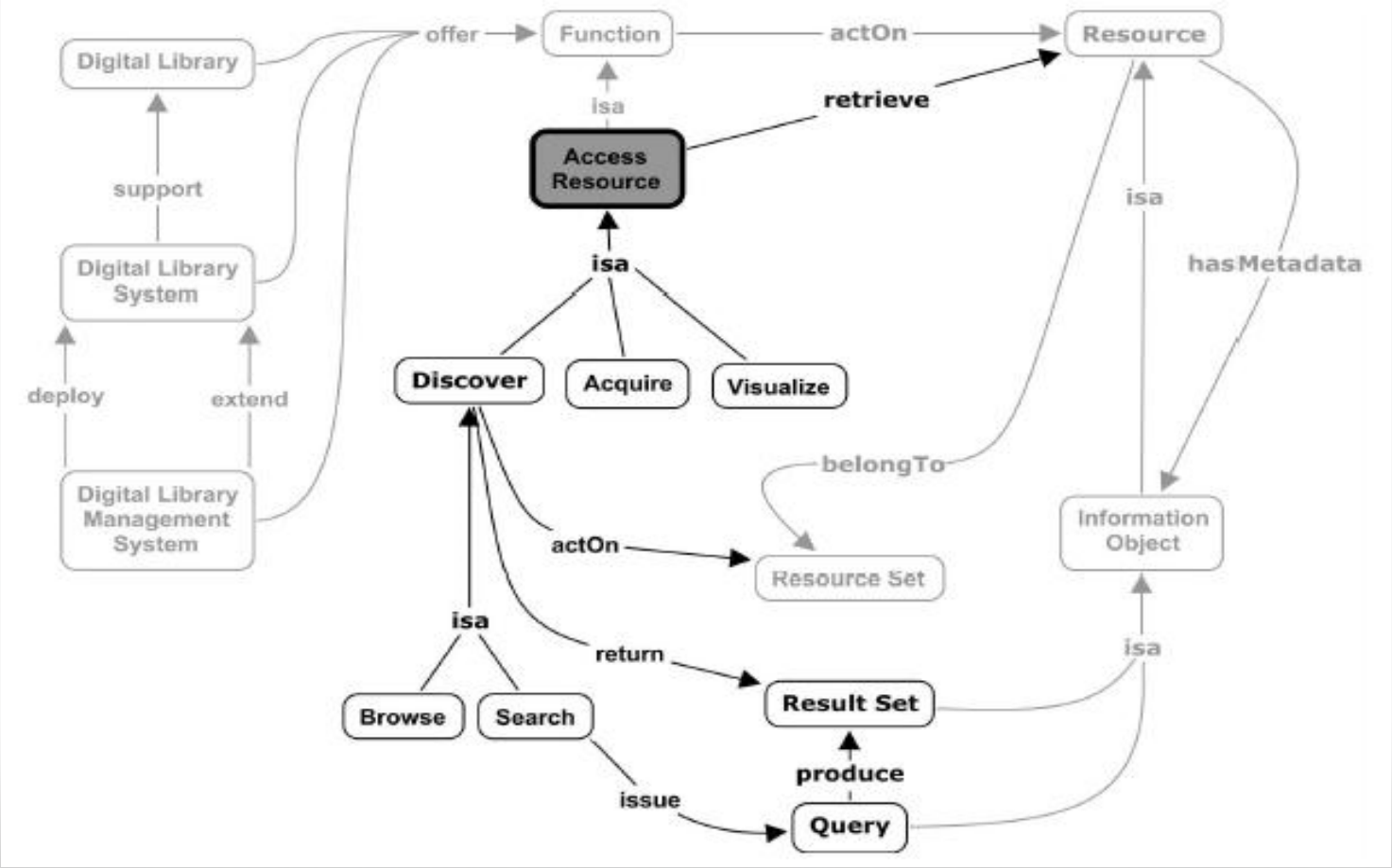
Function is a **Resource**

Actors perform **Functions**

Main **Function** specializations

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

The Functionality Domain(3/11)



The Functionality Domain(4/11)

Access Resource Functions

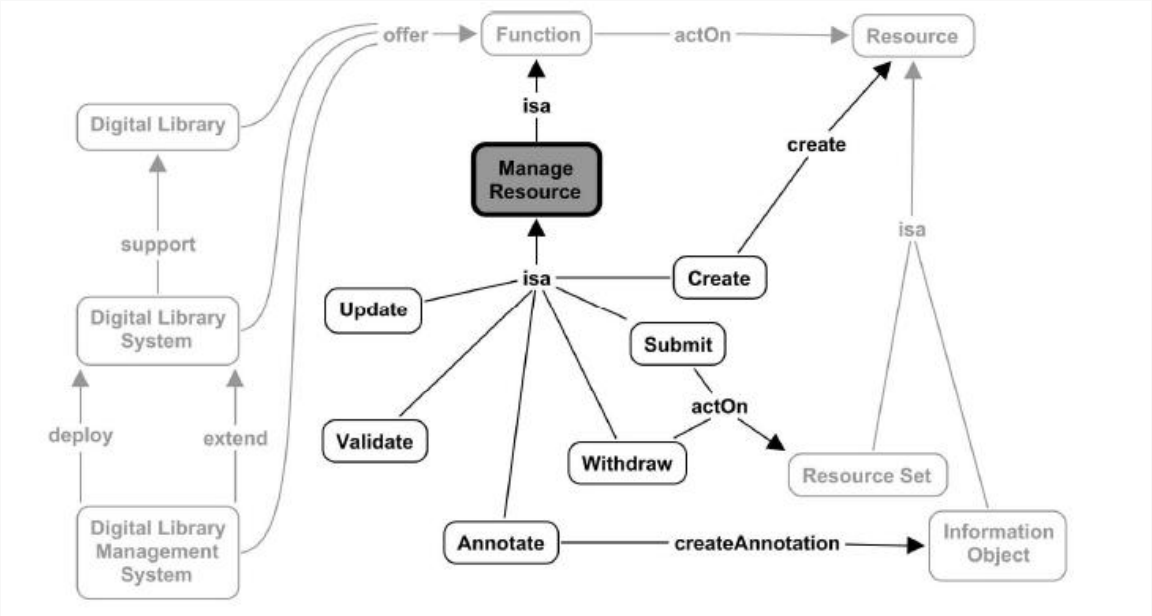
All activities are related to requesting, locating, retrieving and finally delivering *Resources* to requestor

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

The Functionality Domain(5/11)

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

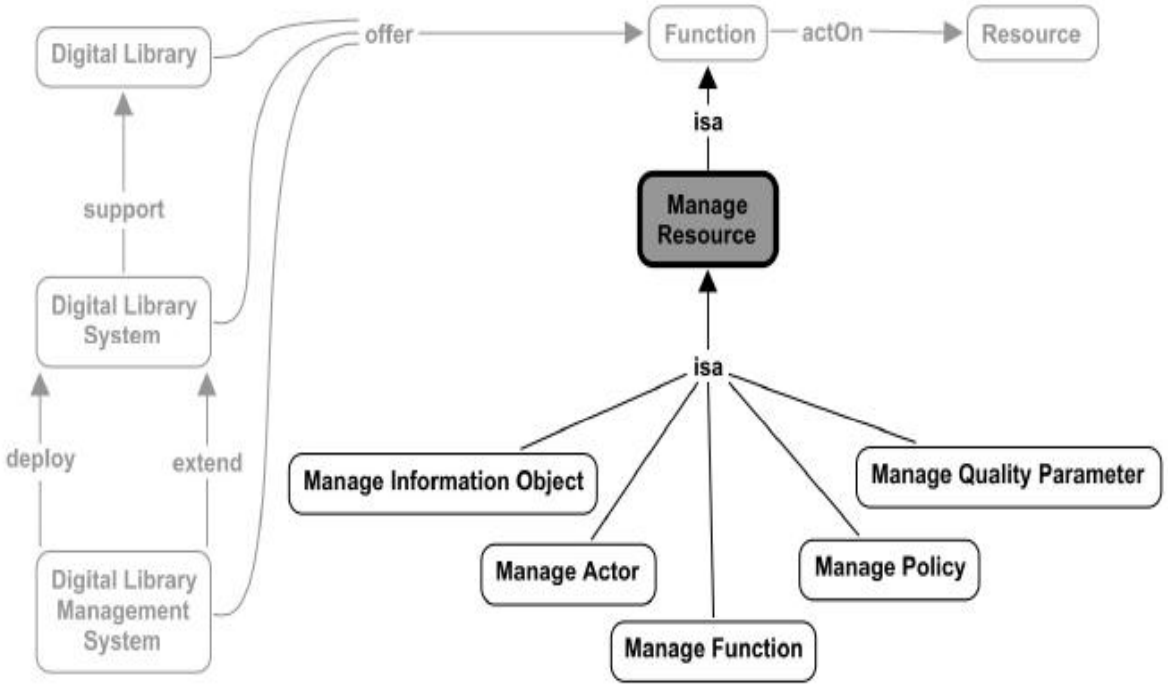


Contains general families of Functions for managing individual Resource Types

The Functionality Domain(6/11)

Manage Resource

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



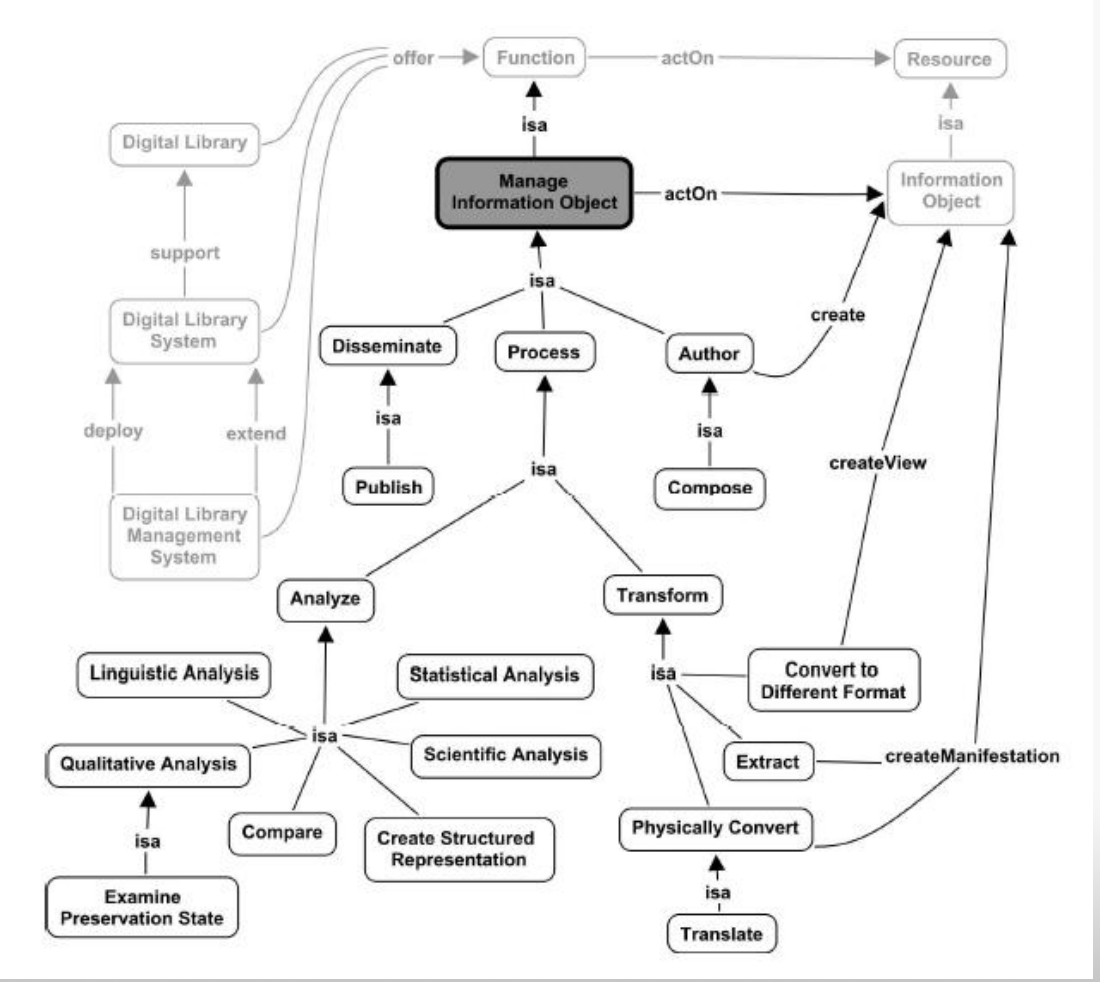
The Functionality Domain(7/11)

Manage Information Object

Captures creation, processing, transformation

primary Information Objects

other Information Objects or Resources in general

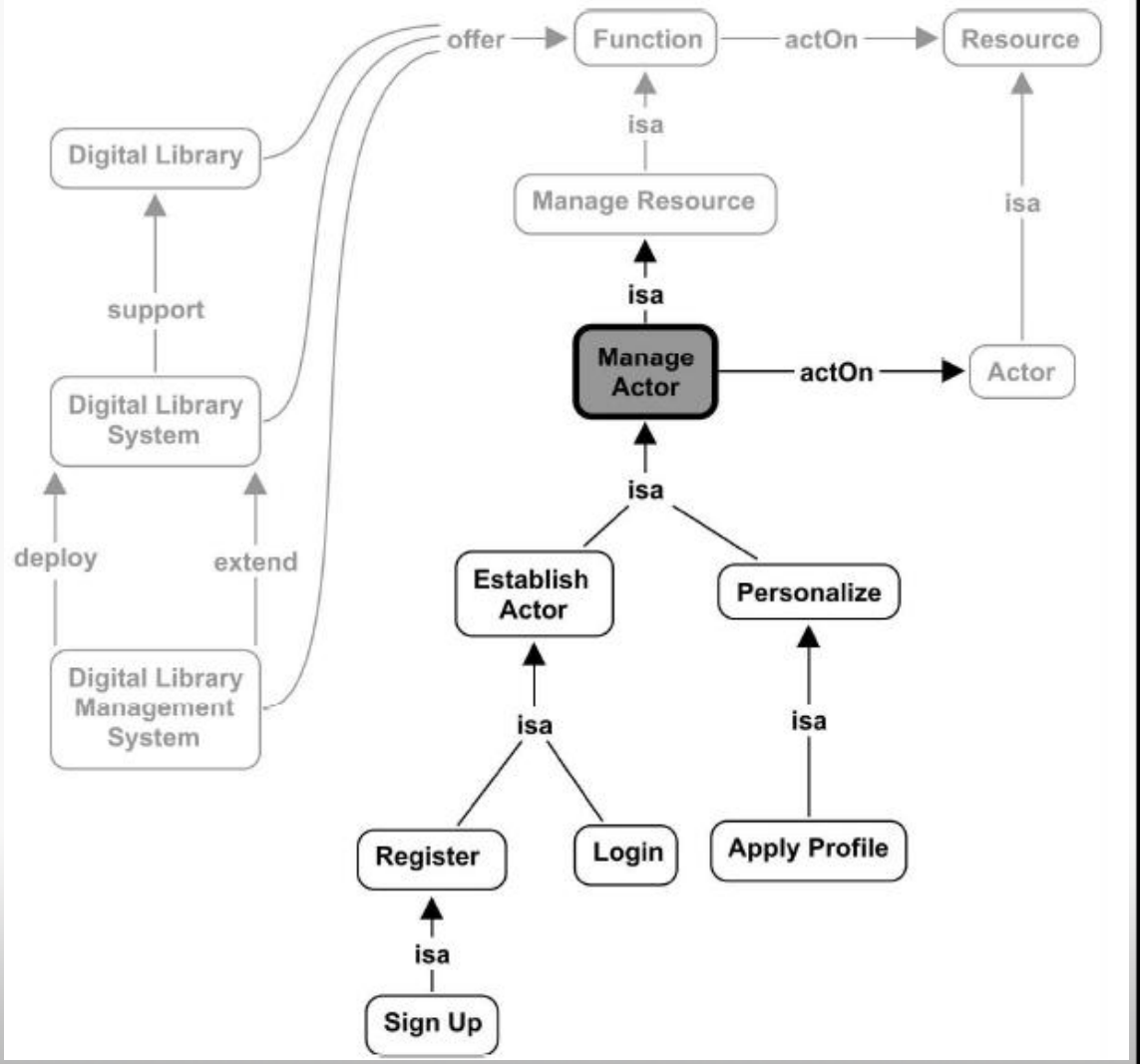


The Functionality Domain(8/11)

Manage Actor

Captures management of individual Actors

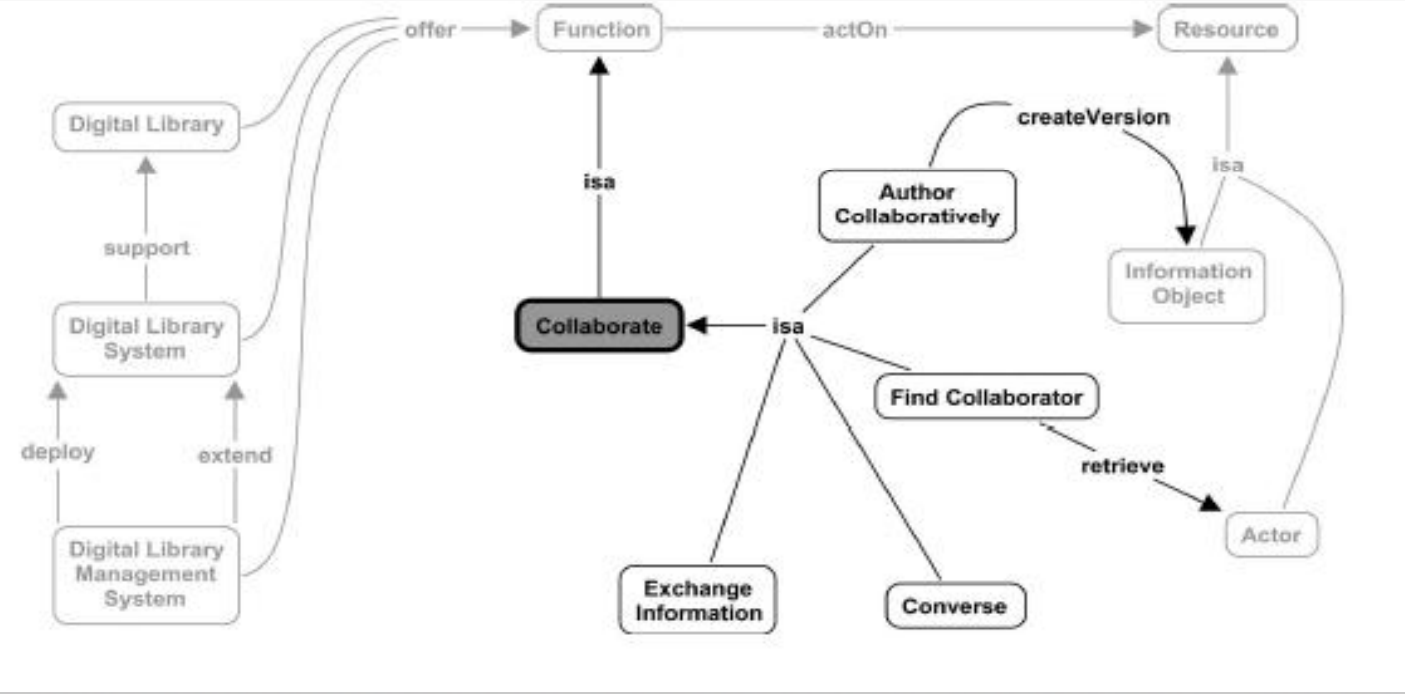
- registration & subscription
- Login
- personalization of allowed functionality



The Functionality Domain(9/11)

Collaborate

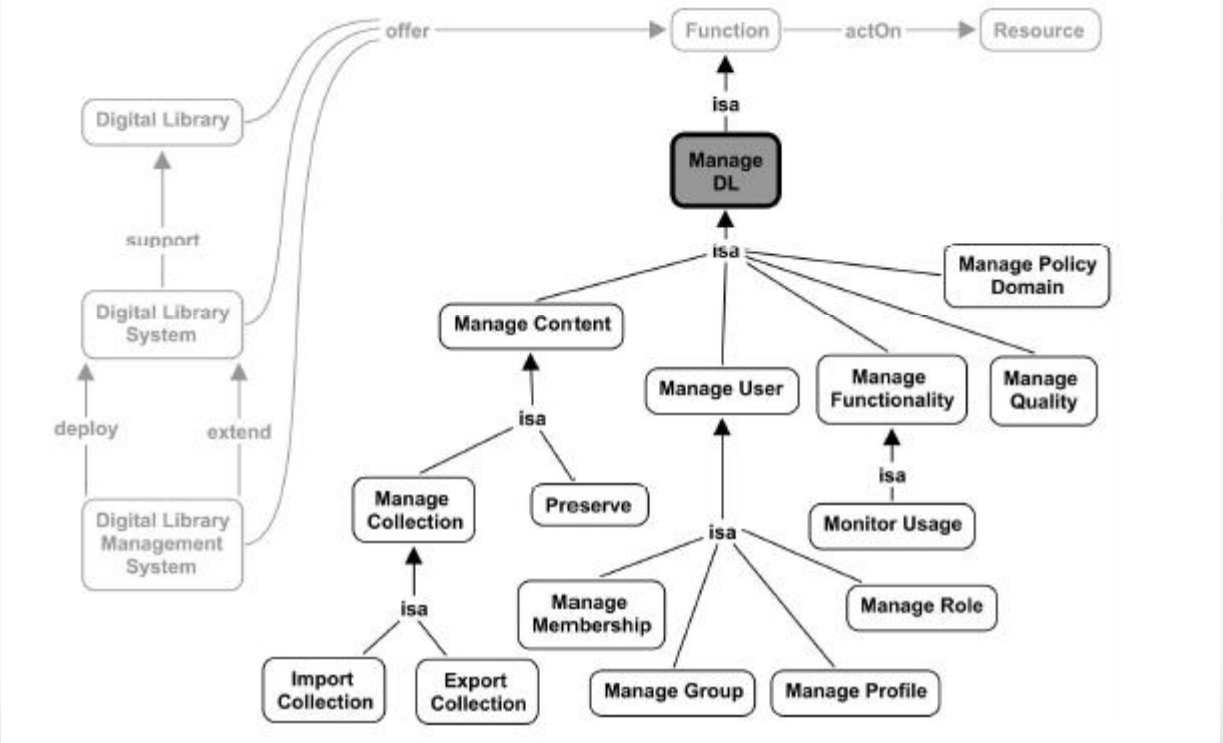
Captures all activities that allow multiple Actors to work together through a DL to achieve a common goal



The Functionality Domain(10/11)

Manage DL

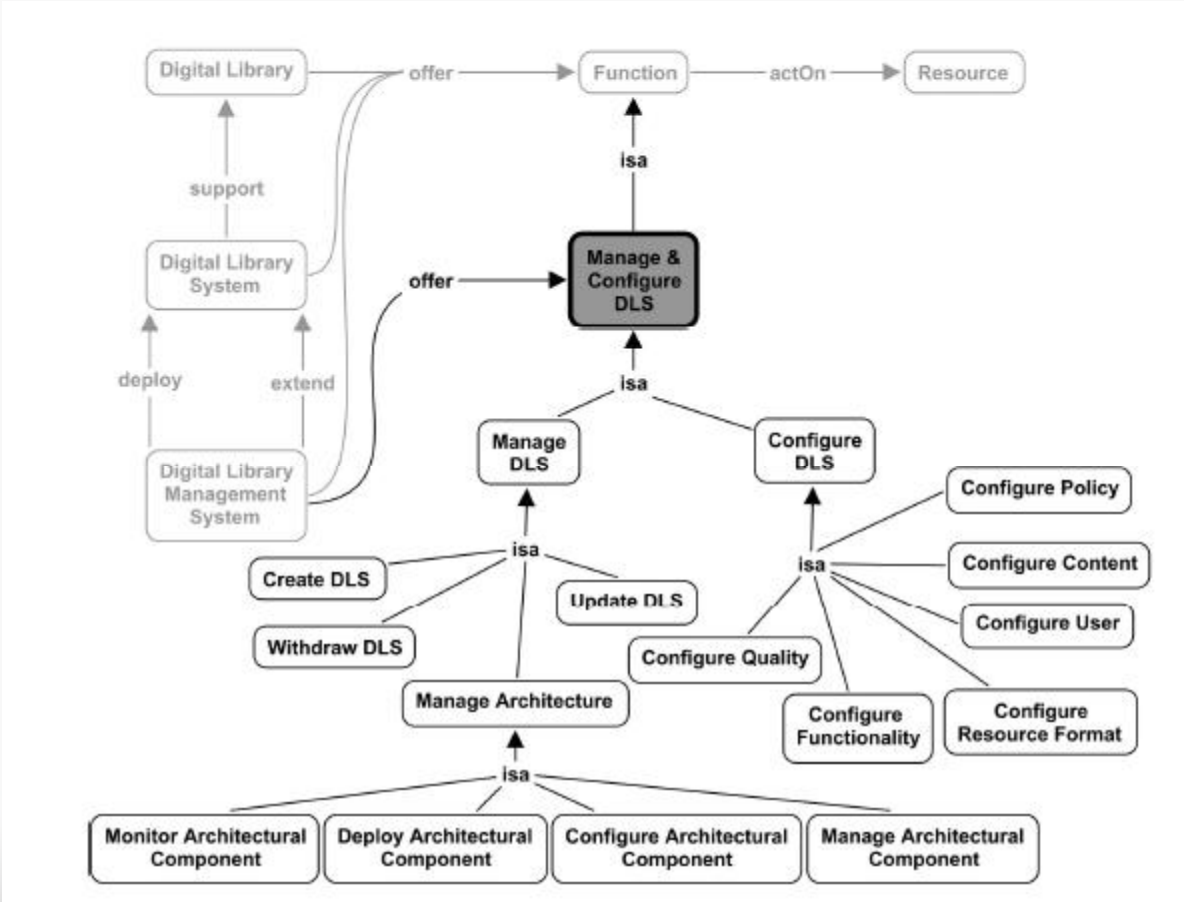
supports the day-by-day DL management, concerning all DL domains



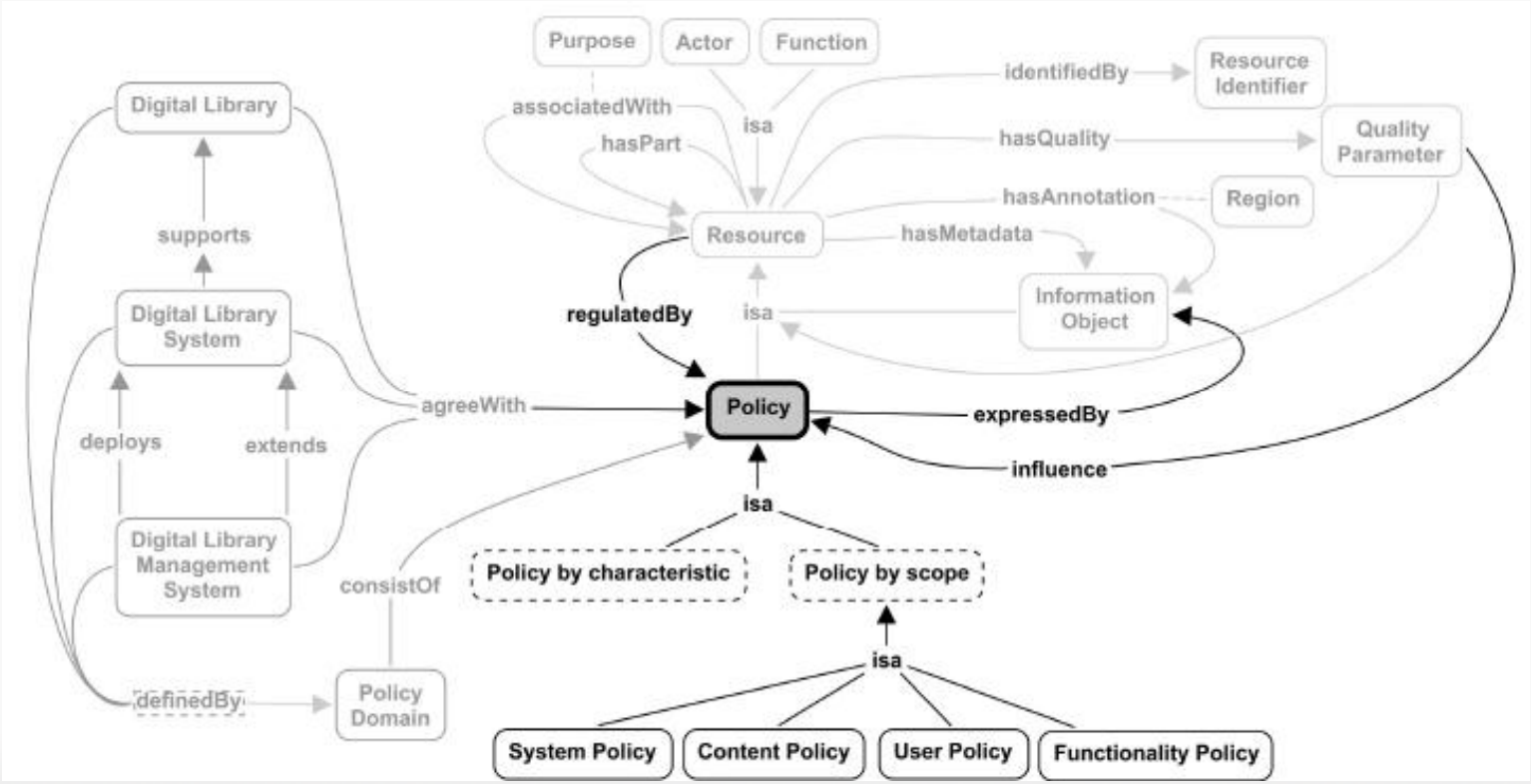
The Functionality Domain(11/11)

Manage and Configure DL

supports setting up, configuring, monitoring the DL



The Policy Domain (1/3)



The Policy Domain (2/3)

Policy Domain:

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**

The Policy Domain (3/3)

Policy by Characteristic:

Extrinsic Policy *vs.* **Intrinsic Policy**

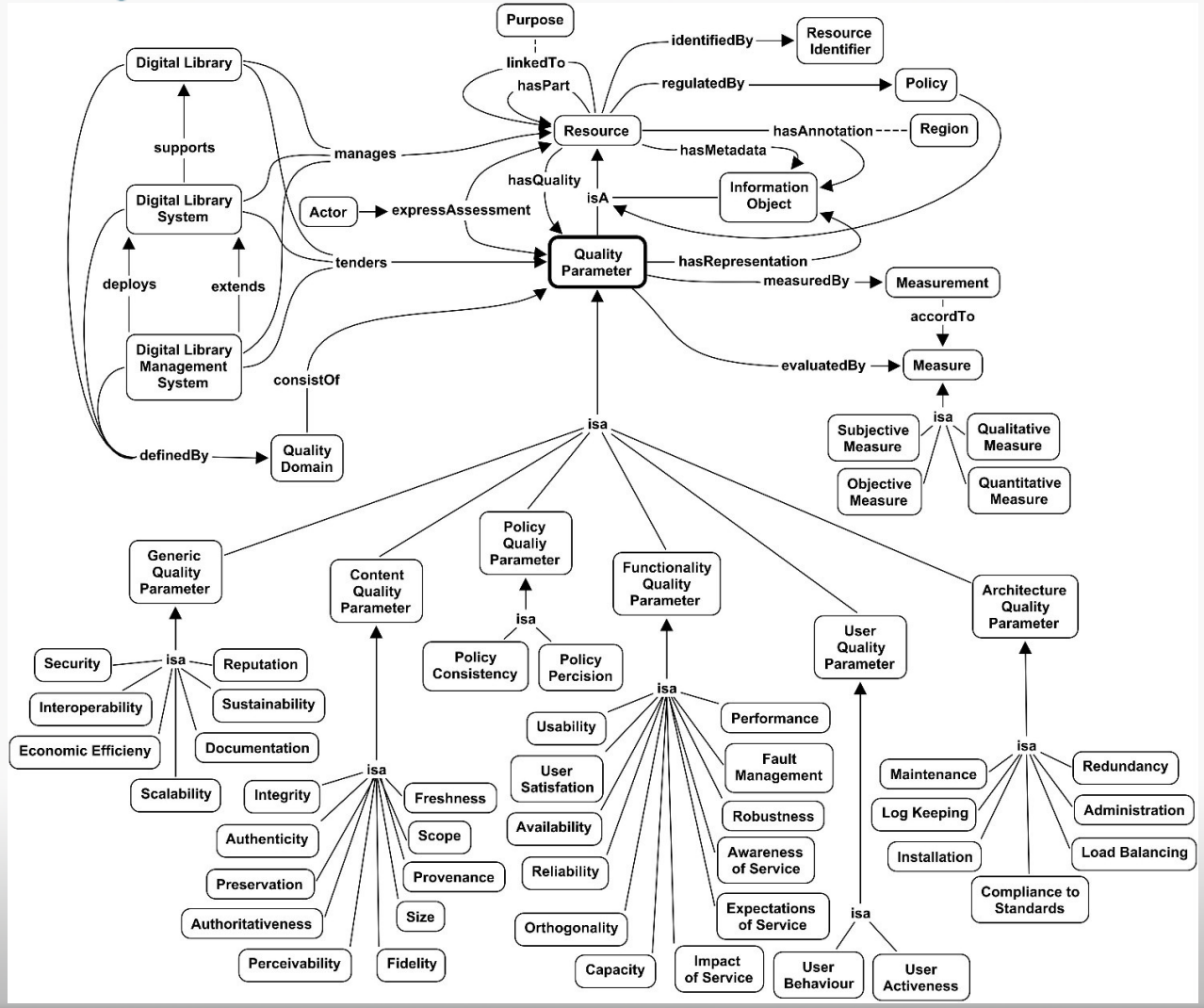
Implicit Policy *vs.* **Explicit Policy**

Prescriptive Policy *vs.* **Descriptive Policy**

Policy by Type

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

The Quality Domain (1/4)



The Quality Domain (2/4)

The Quality Domain:

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**

The Quality Domain (3/4)

Quality Parameter Groups*:*

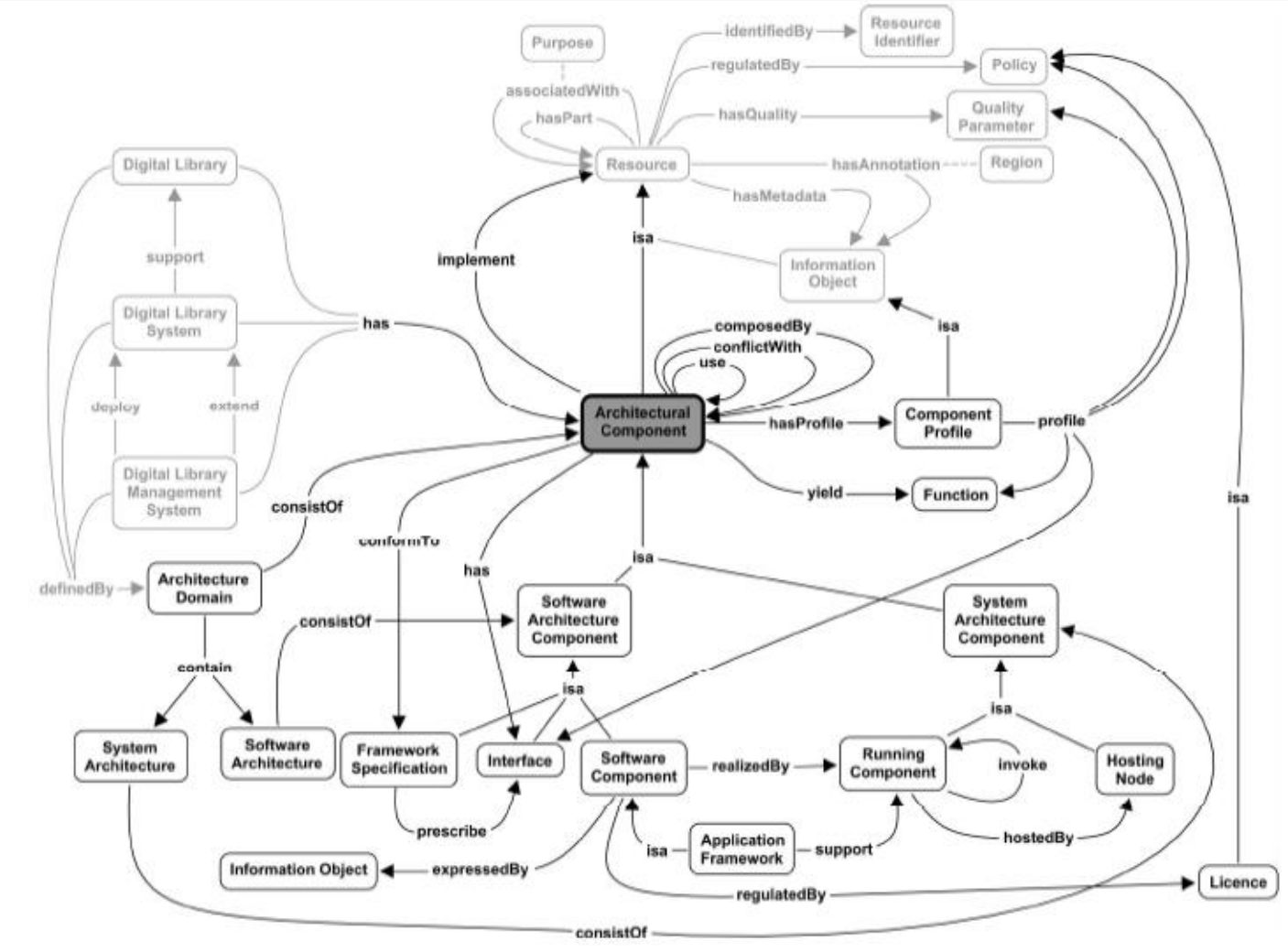
Generic Quality Parameters apply to any kind or most kinds of resources
System Quality Parameters apply to *Digital Library*, or a *Digital Library System*, or a *Digital Library Management System*.
Content Quality Parameters apply to *Resources* in the *Content Domain,* primarily *Information Objects*.
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*

The Architecture Domain (1/3)



The Architecture Domain (2/3)

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 (3/3)

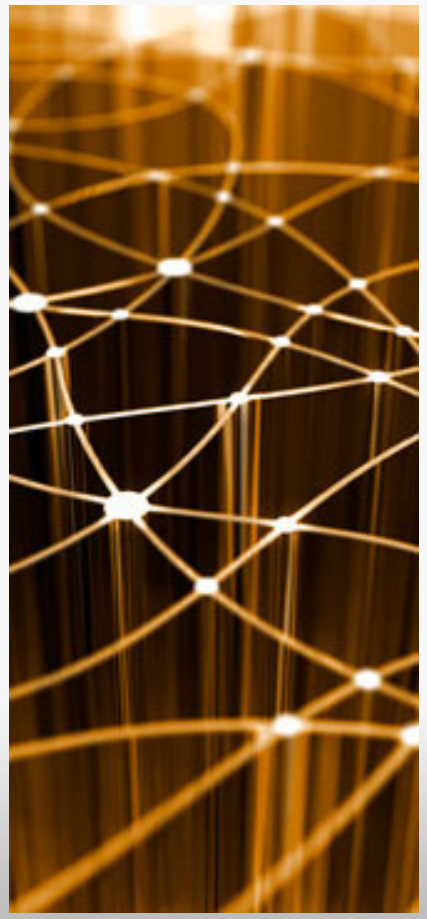
"Component-based approach":

- 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

From DELOS Reference Model to DL.org Outcomes

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

DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations



Thank you