DL.org

Digital Library Interoperability, Best Practices and Modelling Foundations

User Working Group
Towards User Interoperability

Yannis Ioannidis
University of Athens, Greece

1st DL.org Workshop @ECDL 2009, Corfu, Greece, 1 October 2009

Working Group Members

Tiziana Catarci   University of Rome "La Sapienza", Italy

Yannis Ioannidis   University of Athens, Greece

Akrivi Katifori   University of Athens, Greece

Georgia Koutrika   Stanford University, USA

Natalia Manola   University of Athens, Greece

Andreas Nürnberger   University of Magdeburg, Germany

Paul Polydoras   University of Athens, Greece

Manfred Thaller   University of Cologne, Germany

Working Group Members

Working Group Objectives

The User working group has a threefold goal:

- Identifying and deliberating the most important interoperability issues that prevent heterogeneous DL systems from working together from the User perspective.

- Discussing the state-of-the-art regarding implementations that resolve the interoperability issues identified.

- Proposing patterns of approaches that are effective in such a resolution.

The User Domain

- **Actors** entitled to interact with Digital Libraries
- Umbrella concept for all notions related to the representation & management of actors within a DL
- An "actor" can be:
  - individual person
  - group of people acting in unison
  - inanimate entities, e.g., software programs, instruments

The User Domain
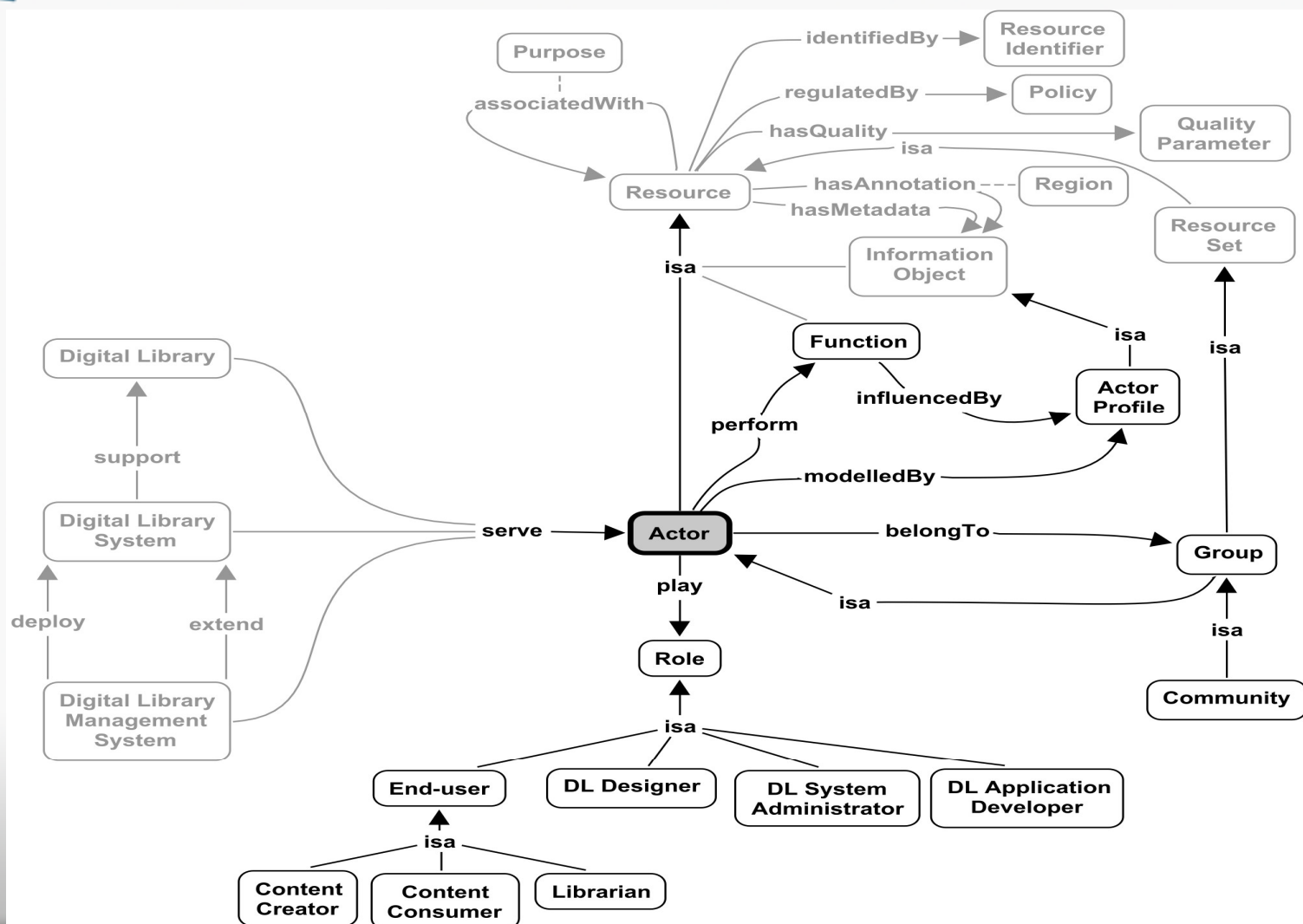
The Main User Roles

- End-users

- DL designers

- DL system administrators

- DL application developers

DL Interoperability

- Distributed Heterogeneous Digital Libraries
- Information in all forms
- RM a future unifying factor, but *interoperability* crucial for
  - legacy systems
  - reconciling different future approaches
- Why/when: integration, composition, matching, mapping, deduction, and activation

Interoperability Abstraction Levels

- Superficial
  - Common tools and interfaces for navigation & access
  - Human intelligence for content coherence
- Syntactic
  - Common metadata models and object transmission protocols and formats for limited coherence
  - Supplementary human interpretation
- Semantic
  - Consistent and semantically coherent access to all digital objects and services
  - Federating/mediating software for site-by-site variations
  - "No" human involvement

"Artists"'s Rendition of DL (Sys)

Policy

Quality

Functionality

User

Content
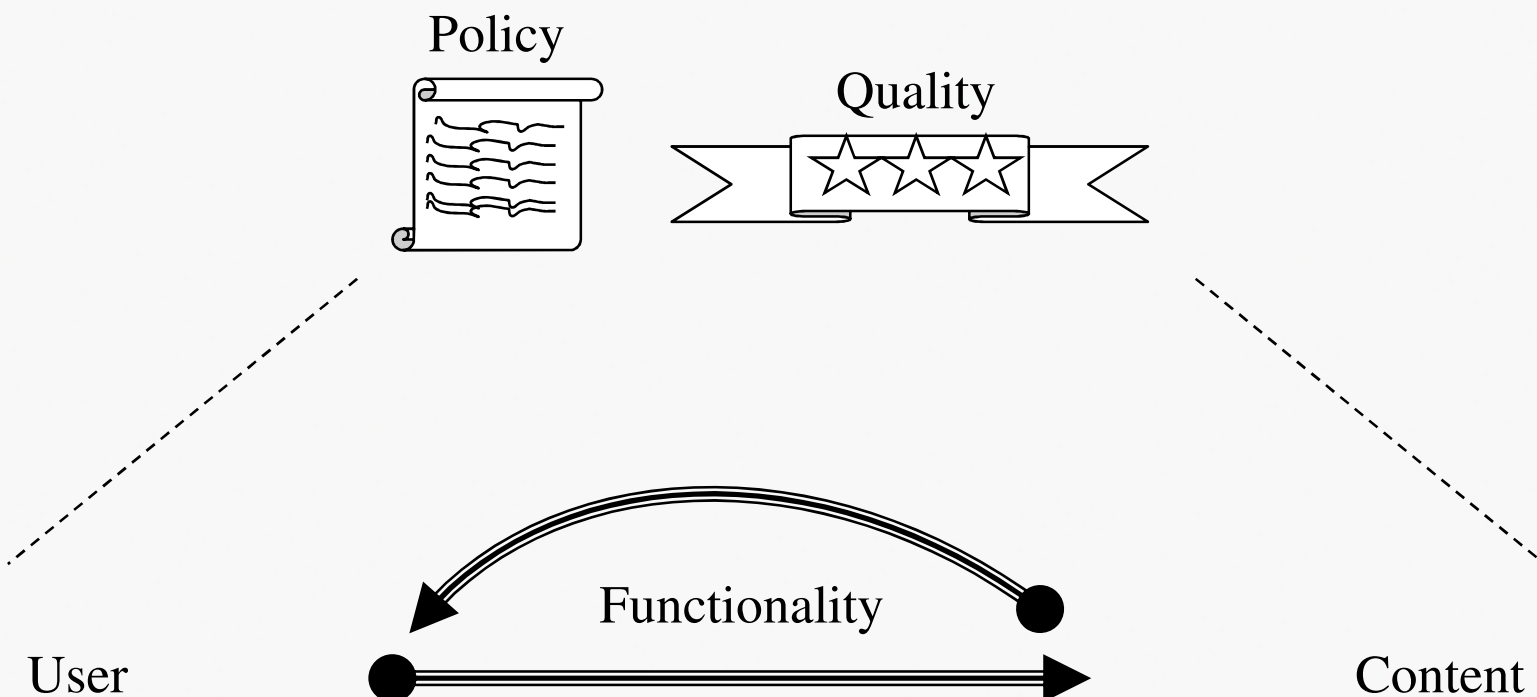
Architecture
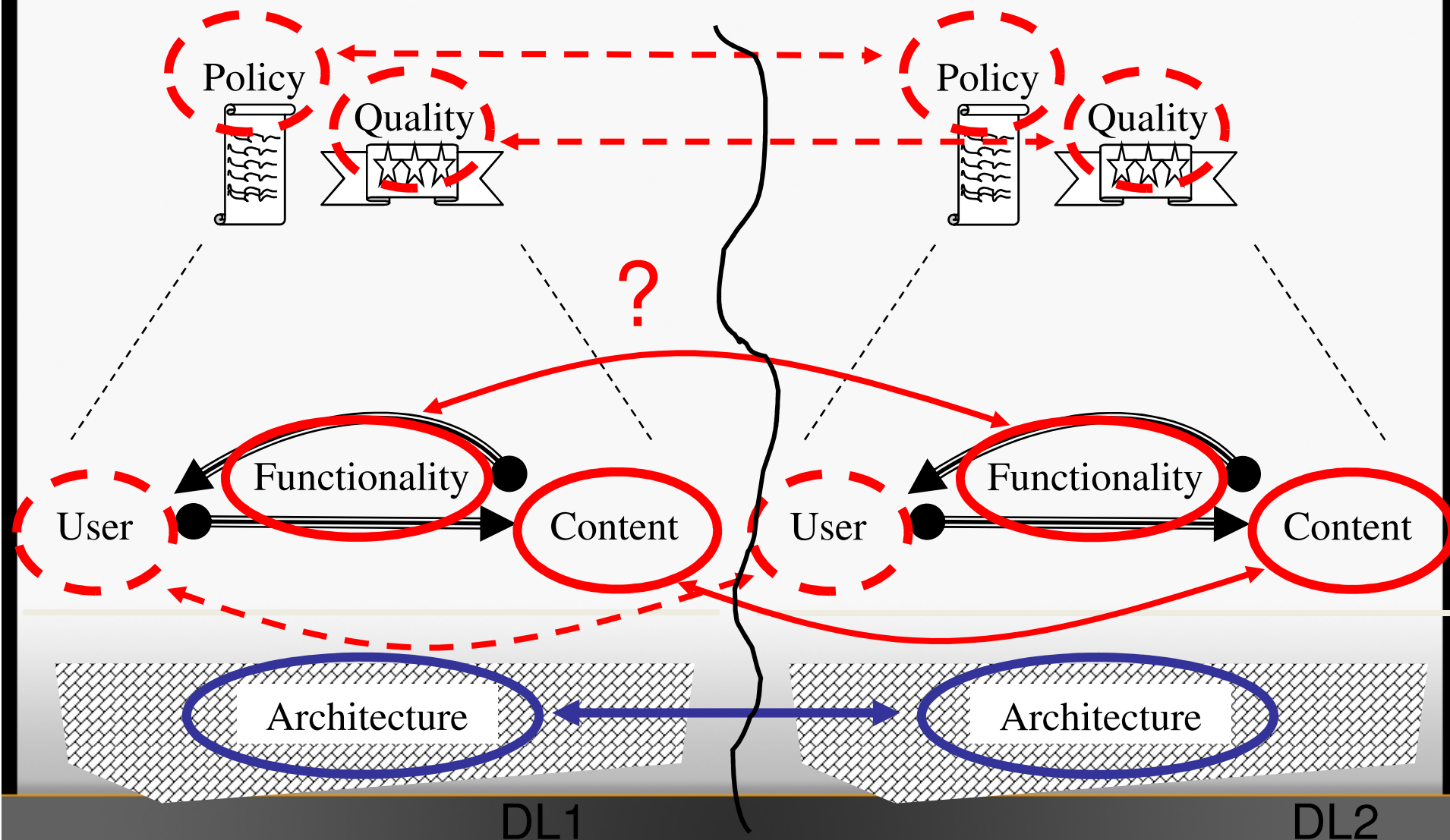
User Interoperability Example

- Donatella@DL1:
  - "Research Infrastructures" $\rightarrow$ 0.9
  - "Swimming" $\rightarrow$ 0.3

- Donatella@DL2:
  - "Research Infrastructures" $\leq$ "Swimming"

- Contradicting or Incomparable?

- Context dependent?

- Reconciliation approach?
  - E.g., More info and stronger statement in DL1

User Interoperability

- Interoperability of DLs/DLSs with regard to what is captured within each DL/DLS about users

- Interoperability of users through their use of the DL/DLS

User Interoperability Scenarios

- Collect, exchange, and integrate information on users: profiles, preferences, access rights, …

- User migration across systems

- Local or distributed operation

- Same services and system behaviour

- User collaboration

Interoperability of DLs/DLSs with respect to users (1/3)

- The "object" of interoperation
  - can be arbitrary
  - can be an attribute of the user (e.g., user credentials, user demographics)
  - can be simple (e.g., keywords)
  - or complex in structure (e.g. ontologies, queries, layouts)
  - can be at the data or at the model/schema level

Interoperability of DLs/DLSs with respect to users (2/3)

- The "purpose" of interoperation
  - preserving user characteristics across systems (transparent user mobility from one system to the next)
  - mapping user characteristics from one system to the next (non-transparent user mobility)
  - integrating user characteristics maintained about the same user in two different systems

- Use cases by combining "objects" and "purposes"
  - consolidating a user's preferences as perceived from his/her presence in multiple systems
  - retaining the user's access rights as the system transfers him/her to another system
  - …

Interoperability of Users

- Through the DL, users are able to
  - collaborate
  - communicate
  - cooperate
- The DLS supports them in
  - knowledge sharing
  - sense making
  - identifying new and/or hidden semantics
- The DLS preserves user privacy and generates a sense of trust

Scope of the Working Group - Interoperability Issues

- Interoperability of DLs/DLSs with respect to users
  - User modeling
  - User profiling (including privacy issues)
  - User context
  - User management
- Interoperability of users
  - Collaboration
  - Participation
  - Privacy

User Modeling Issue (1/4)

- User model captures the essential kinds of info for adaptive system behavior depending on the user

- Attributes of the User that could be reflected in a DL
  - user credentials
  - user demographics
  - user access rights
  - user preferences
  - user interests
  - user background
  - user level of maturity and expertise
  - …

- Users are "entities" with model-based profiles for
  - different access to content (rights)
  - different access to system functionalities (roles)
  - for explicit or implicit preferences affecting the results of user operations
  - for differentiating based on the user context
- A user model of a DL should be rich enough to capture these aspects

User Modeling Issue (3/4)

- Up to now, no generally accepted user model.

- Potential solution: mapping mechanisms within DLs between different user models

- Representation of user models
  - Non-ontological representations
    - relational database
    - XML-based language
  - Ontologies
    - increase the probability that user characteristics will be shared among a range of systems

Interoperable User Models (1/2)

- UserML – User Modeling Markup Language
  - XML-based exchange language based on an ontology that defines the semantics of the XML vocabulary
  - platform for communication about partial user models

- GUMO – General User Model Ontology
  - divide descriptions of user model dimensions into three parts: auxiliary - predicate – range
  - key feature: semantics for all user model and context dimensions mapped to general ontology

- GUC – Generic User model Component
  - generic component w/ functionality to store data models for applications and to exchange user data between them

- SUMI – Scrutable User Modeling Infrastructure
  - A model is an integration of various user models (obtained by interacting with various services on the WWW)
  - Users able to export part of their SUMI model to any registered service they prefer (SUMI export protocol, based on Semantic Web)

User Profiling Issue (1/3)

- Process of collecting information about a user to generate their profile, based on current user model
- Interoperable DL systems offer personalized DL usage experience
- Challenges
  - user rights propagation from one DL to the other
  - reconciliation of different and, in some cases, even conflicting preferences or user profile characteristics
  - Information can be stored in different data structures
    - Flat, Hierarchical, Graph-based, Semantic profiles

User Profiling Issue

- User profiling methods:
  - logging user behaviour and analyzing log files and related objects/ressources (using statistical and machine learning approaches) to derive user characteristics
- Information can be stored in different data structures
  - Flat profiles
  - Hierarchical profiles
  - Graph based profiles
  - Semantic profiles

- Profile acquisition (explicit)
  - User registration
  - User states search "objective"/search keywords at beginning of a session (information need)
  - Explicit relevance feedback
- Profile acquisition (implicit)
  - Log more general user (-system) interaction
  - Implicit relevance feedback
  - Mining of log files to obtain "higher level" properties

User Profiling Issue (3/3)

- Profile extension
  - Group profiles used to derive additional (possibly relevant) information for a specific user
    - whether or not the user is a group member
  - Specific user profiles used to derive information about group profiles

User Context Issue (1/2)

- User context represents "external" factors affecting user profiles regarding user interactions with a DL
- Borders of "external" and "internal" factors hazy
- Context may include the user
  - situation
  - location
  - time
  - role
  - presence of other users
  - ... any other RM domain

User Context Issue (2/2)

- Narrow technical context (OS, top-level UI, …)
- Wide technical context (info environment w/ user help)
- Administrative context (info env w/o user help)
- Semantic context (content-related preferences)

User Context Issue

- **Definition**: The concept of the *context* of a user - or *user context* - covers all properties of an information environment, which are expected to be implicitly available when the user interacts with any component of such an information environment.

- Narrow technical context

- Wide technical context

- Administrative context

- Semantic context

User Context Issue

- **Definition**: The *narrow technical context* of a DL describes all settings of an operating system and / or top level user interface, which describe preferences of a human user or of a program interacting with arbitrary software components.

- Relationships: "Narrow technical context <isA> User context"

- **Definition**: The *wide technical context* of a DL describes all components of an information environment, from which the DL can receive hints how to respond to request of a user within a specific situation.

- Relationships: "Wide technical context <isA> User context"

User Context Issue

- **Definition**: The *administrative context* of a DL describes all components of an information environment, which allow a DL to determine the C12 Actor Profile of a user, without the actor explicitly negotiating with the DL about it.

- Relationships: "Administrative context <isA> User context"

- **Definition**: The *semantic context* of a DL describes all content-related preferences which are connected with a user.

- Relationships: "Semantic context <isA> User context"

User Management Issue (1/2)

- User privileges, authentication and authorization functions

- Interoperability: DLSs working in synergy over concrete and shared but user-transparent policies on the above

- User management systems manage electronic identities, thus acting as IDentity Management Systems (IDM).

- Identity management has three perspectives
  - Pure identity paradigm: Creation, management & deletion of identities without regard to access or entitlements
  - User access (log-on) paradigm: For example: a smart card and its associated data used to log on to a service
  - Service paradigm: A system that delivers personalized, role-based, online, on-demand, multimedia (content), presence-based services to users and their devices

Interoperable User Mgmt (1/2)

- Federated identity
  - portability of id info across autonomous security domains
- Open industry standards or openly published specifications for common use cases
- Typical use-cases: cross-domain ...
  - ... web-based single sign-on
  - ... user account provisioning,
  - ... entitlement management
  - ... user attribute exchange.

- Identity federation accomplished in several ways
  - OpenID
  - Security Assertion Markup Language - SAML
  - eXtensible Access Control Markup Language - XACML
  - Liberty Alliance Project Identity Federation Framework – Liberty ID-FF
  - Shibboleth
  - Athens
  - WS-Federation

- Most prominent access management and identity federation systems: PERMIS and Sun OpenSSO

Scope of the Working Group - Interoperability Issues

- Interoperability of DLs/DLSs with respect to users
  - User modeling
  - User profiling (including privacy issues)
  - User context
  - User management
- Interoperability of users
  - Collaboration
  - Participation
  - Privacy

Collaboration Issue (1/3)

- Content should be available to
  - single users
  - cooperating user groups or communities

- Content exchange between users must be
  - simple
  - Intuitive
  - transparent

- DLs: from simple content providers to "platforms" for creative work and production of new knowledge

Collaboration Issue

- Presently, the Digital Libraries efforts are moving beyond the work of simply gathering, curating and providing access of the digital content and look into ways of providing new added value services to their users with a shift in collaboration environments.

- The basic idea behind collaboration is that users/researchers want to exchange information, ideas and views.

Collaboration Issue (2/3)

- Two types of collaboration:
  - indirect
  - direct

- Indirect (passive): work of one user may somehow benefit anonymously from the work (actions) of other users

- Direct: several users agree to work together as a team exploring and making use of DL resources

- Trust and privacy play an important role

Collaboration Issue (3/3)

- The most frequently mechanisms used for indirect collaboration are:
  - Collaborative-based filtering
  - Processing of usage statistics and the use of recommendations as a mean of collaboration.
  - Annotations: users may add content which is complementary to the existing information of a digital object represented in the library and thus share ideas.
  - Tags: this may be viewed as a light form of annotation and it is used as a method to categorize objects.
  - Rankings: a user oriented operation which allows users to share their opinion on a given object.
  - Collection of digital objects: organize information space according to their own subjective perspective
  - Users providing links amongst digital object.

Collaboration Issue

- Direct collaboration takes place in a system where several users agree to work together as a team exploring and making use of digital library resources.

- In more advanced (emerging) situations users working actively together toward a common goal may want to share not only knowledge or information but also use collaborative tools to create new content (shared repository) or act on existing one.

- Collaborative tools are software environments that support various forms of interaction among people.

- Trust and privacy play an important role.

Participation Issue (1/2)

- Making content more "available" & attractive to users

- Not easy allowing users to be at the same time
  - content consumers
  - content providers (in some sense)

- Participation objective
  - not only to support collaboration when users work on the same content
  - but transform DL interaction into an interactive and attractive experience

Participation Issue (2/2)

- Appropriate functionality
  - annotation services
  - translations or transcriptions and other, more active ways to contribute to the DL content

- Crucial issues
  - moderation
  - approval of changes
  - reconciliation
  - provision of different user views on the same content

![DL.org logo — Digital Library Interoperability, Best Practices and Modelling Foundations]

Interoperable User Participation

- Social networks

- Flickr Commons (partner w/ Library Congress + 15 other institutions)

- Noosphere serving as the PlanetMath project's software platform

- Several types of privacy
  - (static and dynamic) privacy of the users accessing the DL
  - privacy of the DL digital objects, that may depend on the context of usage, the purpose, who is requesting the objects, etc.
  - privacy of data, due to
    - multimedia nature of data
    - presence of annotations
    - possible laws and regulations certain objects are subject to

Privacy Issue (2/2)

- Archiving Users' Data and Users' Privacy

- DLs as Online Communities

- Copyright and Privacy

Interoperable User Privacy (1/2)

- SemWebDL - multi-library interoperable setting.

- Two key ideas
  - Definition of 3-tier privacy model for DLs
    - user privacy
    - service privacy
    - library privacy
  - Reputation-based service layer for DL user access

- In SemWebDL the user privacy profile has two components:
  - a static privacy profile
  - a service access profile
- The service privacy policy has two components
  - a user interaction policy
  - a service interaction policy
- The library privacy policy has two components
  - an access policy
  - a data privacy policy