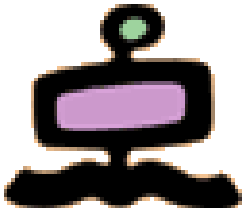
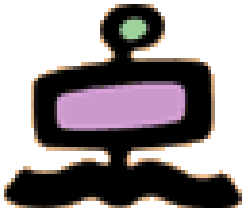
Corso di Biblioteche Digitali



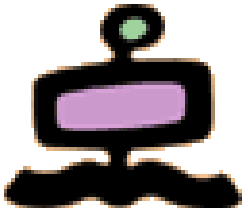
- Vittore Casarosa
  - Mail: casarosa@isti.cnr.it
  - Tel: 050 621 3115 (office) 348 397 2168 (mobile)
  - Skype: vittore1201
- “Ricevimento” at the end of the lessons or by appointment
- Final assessment
  - 70% oral examination
  - 30% project (development of a small digital library))
- Reference material:
  - Ian Witten, David Bainbridge, David Nichols, How to build a Digital Library, Morgan Kaufmann, 2010, ISBN 978-0-12-374857-7 (Second edition)
  - Material provided by the teacher
- **http://cloudone.isti.cnr.it/casarosa/BDG/**

Modules



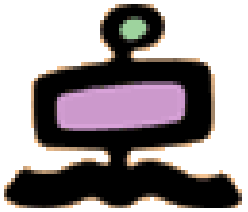
- Computer Fundamentals and Networking
- A conceptual model for Digital Libraries
- Bibliographic records and metadata
- Information Retrieval and Search Engines
- Knowledge representation
- Digital Libraries and the Web
- Hands-on laboratory: the Greenstone system

The role of libraries



- Centuries and centuries of history
- Mediators between information and users
- Selection
  - Definition of collections
- Acquisition
  - Physical objects
- Description
  - Catalogs
- Access
  - Shelves
- Preservation
  - Controlled enviroment

Libraries: some (old) figures

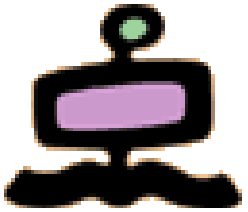


- Volumes (in millions)

Institution	1910	1995	2002	2012
Library of Congress	1,8	23	26	34
Harvard Univ.	0,8	12,9	14,9	16
Yale Univ.	0,55	9,5	10,9	12
U Illinois (Urbana)	0,1	8,5	9,9	11
U California (Berkeley)	0,24	8,1	9,4	10
British Library	2	15	18	25
Cambridge Univ.	0,5	3,5	7	
Oxford Univ.	0,8	4,8	6	
Bibl. Nat. De France	3	11	12	

- Journals
  - From 10.000 in 1950 to 150.000 in 2002
- Alexandria principle beginning to fade

Library of Congress in 2022

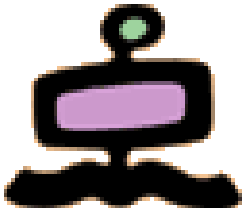


Recorded more than 175.77 million items in its collections, including:

- more than 25,5 million cataloged books in the Library of Congress classification system
- more than 15,9 million items in the non-classified print collections (incunabula, monographs and serials, music, bound newspapers, pamphlets, technical reports , etc)
- more than 134.4 million items in the non-classified special collections, including:
  - 4.2 million audio materials (discs, tapes, talking books and other recorded formats)
  - 77 million manuscripts
  - 5.8 million maps
  - 17.5 million microforms
  - 1.8 million moving images
  - 8.2 million items of sheet music
  - 16 million visual materials
  - 2 million other items (including machine-readable items)

3172 staff members
840 millions $

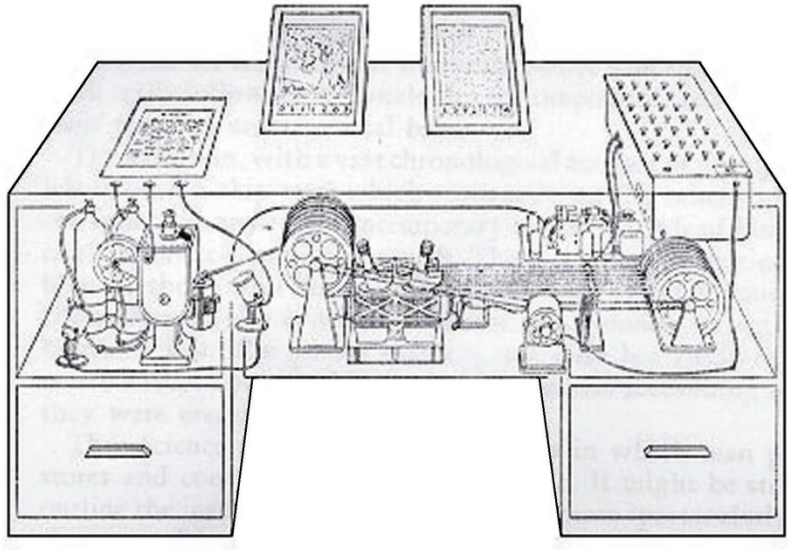
Use of computers in the libraries



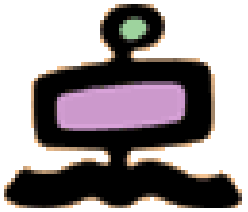
- Early technodreams
  - Vannevar Bush (1890-1974)
    - Paper in 1945: As we may think
  - JCR (Joseph Carl Robnett) Licklider (1915-1990)
    - Book in 1965: Libraries of the future

Vannevar Bush
As we may think - 1945

- Head of US science during WW2
- Use of “knowledge” and team work to advance Science
- The Memex: mechanized private archive and library (microfilms)
- “trails” of information
  - associative links (roughly the equivalent of hyperlinks)
- No “free text” search

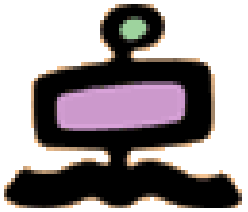


JCR Licklider
Libraries of the future - 1965



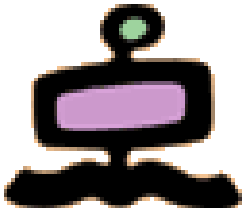
- Head of US Dept. of Defense, Information Processing Technologies (supporter of DARPANET)
- The book foresees the research and development needed to build a Digital Library
  - Time-sharing just beginning
  - “Big” memories around 32K
  - Networking “to be invented”
- Rather accurate overall view of what a DL could look like in 1995
  - Under-estimation of computing power
  - Over-estimation of progress in
    - Artificial intelligence
    - Natural language processing

What are Digital Libraries ?



- Digital Libraries do exist today
  - Are they a transformation of “traditional libraries”?
  - Are they an evolution of data bases ?
  - Are they (a subset of) the Web ?
  - Are they useful ?
- DLs are at the intersection of a number of different disciplines/technologies
- A “theory” of Digital Libraries not yet developed
- Two perspectives
  - The Digital Library Curriculum Project
  - The DELOS Reference Model

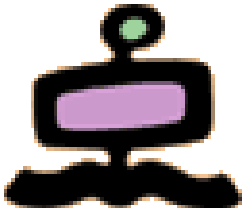
Digital Library Curriculum Project



Bottom-up approach to identify major topics related to Digital Libraries in about 65 university courses (2005-2009), identifying 43 topics, divided in 10 “subjects”

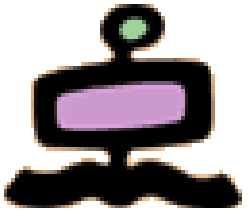
- 1 - Overview
- 2 - Digital Objects
- 3 - Collection Development
- 4 - Info/Knowledge organization
- 5 - Architectures (agents, mediators)
- 6 - User Behavior/Interactions
- 7 - Services
- 8 - Preservation
- 9 - Management and Evaluation
- 10 - DL education and research

Topics underlying DLs (1)



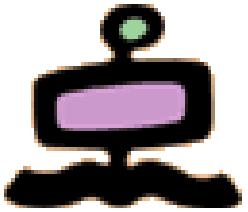
- 1 - Overview
  - 1-a (10-c): Conceptual frameworks, theories, definitions
  - 1-b: History of digital libraries and library automation
- 2 - Digital Objects
  - 2-a: Text resources
  - 2-b: Multimedia
  - 2-c (8-c): File formats, transformation, migration
- 3 - Collection Development
  - 3-a: Collection development/selection policies
  - 3-b: Digitization
  - 3-c: Harvesting
  - 3-d: Document and e-publishing/presentation markup
- 4 - Info/Knowledge organization
  - 4-a: Information architecture (e.g., hypertext, hypermedia)
  - 4-b: Metadata, cataloging, metadata markup, metadata harvesting
  - 4-c: Ontologies, classification, categorization
  - 4-d: Subject description, vocabulary control, thesauri, terminologies
  - 4-e: Object description and organization for a specific domain

Topics underlying DLs (2)



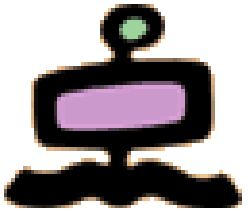
- 5 - Architectures (agents, mediators)
  - 5-a: Architecture overviews/models
  - 5-b: Application software
  - 5-c: Identifiers, handles, DOI, PURL
  - 5-d: Protocols
  - 5-e: Interoperability
  - 5-f: Security
- 6 - User Behavior/Interactions
  - 6-a: Info needs, relevance
  - 6-b: Search strategy, info seeking, behavior, user modeling
  - 6-c: Sharing, networking, interchange (e.g., social)
  - 6-d: Interaction design, info summarization and visualization, usability assessment
- 7 - Services
  - 7-a: Search engines, IR, indexing methods
  - 7-b: Reference services
  - 7-c: Recommender systems
  - 7-d: Routing, community filtering
  - 7-e: Web publishing (e.g., wiki, rss, Moodle, etc.)

Topics underlying DLs (3)



- 8 - Preservation
  - 8-a: Approaches to archiving and repository development
  - 8-b: Sustainability
  - 8-c (2-c): File formats, transformation, migration
- 9 - Management and Evaluation
  - 9-a: Project management
  - 9-b: DL case studies
  - 9-c: DL evaluation, user studies
  - 9-d: Bibliometrics, Webometrics
  - 9-e: Legal issues (e.g., copyright)
  - 9-f: Cost/economic issues
  - 9-g: Social issues
- 10 - DL education and research
  - 10-a: Future of DLs
  - 10-b: Education for digital librarians
  - 10-c (1-a): Conceptual framework, theories, definitions
  - 10-d: DL research initiatives

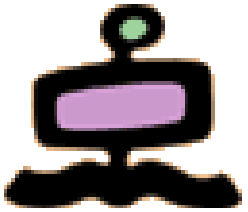
DELOS NoE (2000-2008)



The Network of Excellence on Digital Libraries

Define and conduct a **joint program of activities** in order to integrate and coordinate the on-going research activities of the major European research teams in the field of **digital libraries** for the purpose of developing the next generation digital library technologies

DELOS visions



YEAR 2000

- Digital libraries should enable any citizen to access all human knowledge anytime and anywhere,in a friendly, multi-modal, efficient, and effective way, by overcoming barriers of distance, language, and culture and by using multiple Internet-connected devices

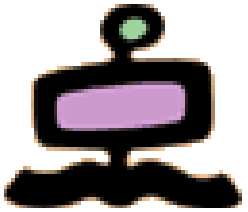
YEAR 2005

- The potential exists for digital libraries to become the universal knowledge repositories and communication conduits for the future, a common vehicle by which everyone will access, discuss, evaluate, and enhance information of all forms

Bibliography of Digital Library definitons

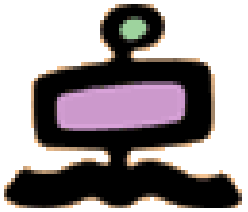
- [1] Kuny, T ; Cleveland, G ‘The Digital Library: Myths and Challenges’ In: Proceedings 62nd IFLA General Conference, August 1996
- [2] Lesk M ‘Practical Digital Libraries, Books, Bytes and bucks’ Morgan Kaufman, 1997
- [3] Belkin, N J ‘Understanding and Supporting Multiple Information Seeking Behaviors in a Single Interface Framework’ In: User Interfaces in Digital Libraries, ERCIM, 1999
- [4] Borgmann C ‘What are Digital Libraries ? Competing visions’ In: Information Processing and Management, 35(3), 1999
- [5] Lesk, M ‘Expanding Digital Library Research: Media, Genre, Place and Subjects’ In: Proceedings of the International Symposium on Digital Libraries, ISDL’99, Tsukuba, Japan, 1999
- [6] ‘Digital Library Policies, Organizations and Practices’ Digital Library Federation Survey, 1999
- [7] Soergel, D ‘A Framework for Digital Library Research’ DLib Magazine, 8(12), December 2002
- [8] ‘Digital Libraries in Education’ UNESCO (IITE) Analytical Survey, 2003
- [9] i2010 Digital Libraries: Communication form the Commission to the European Parliament, Bruxelles, September 2005
- [10] Brahaj A.,Razum M.,Hoxha J., ‘Defining Digital Library’, Proceedings of TPDL 2013, Springer-Verlag, 2013

Digital Library definitions (1)



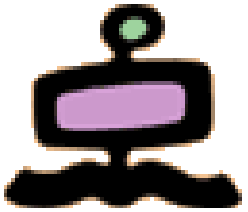
- Digital Libraries impose reinvention of the role of librarians and library models, as none of the following is true: (i) the Internet is 'The' Digital Library; (ii) at some point there will be a single Digital Library or a single-window view of Digital Library collections; (iii) Digital Libraries are means to provide more equitable access to content from anywhere at any time; and (iv) Digital Libraries are cheaper instruments than physical libraries [1]
- Digital Libraries are organized collections of digital information They combine the structure and gathering of information, which libraries and archives have always done, with the digital representation that computers have made possible [2]
- Digital Library is an institution responsible for providing at least the functionality of a traditional library in the context of distributed and networked collections of information objects [3]
- Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching and using information; they are an extension and enhancement of information storage and retrieval systems that manipulate digital data in any medium Digital libraries are constructed, collected and organized by (and for) a community of users, and their functional capabilities support the information needs and uses of that community Researchers view Digital Libraries as content collected on behalf of user communities, while practising librarians view Digital Libraries as institutions or services [4]

Digital Library definitions (2)



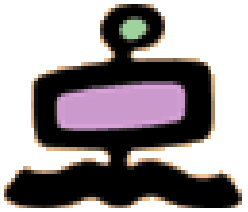
- Digital Libraries are tools to serve research, scholarship and education; tools to access information; tools to provide services primarily to individual users [5]
- 'Digital' mainly implies the existence of software for searching text, while "Library" mainly refers to existing material that has been scanned for online access (in both cases research efforts not usually associated with users' needs) [6]
- Digital Libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works, so that they are readily and economically available for use by a defined community or set of communities [7]
- Digital libraries are organized collections of information resources in digital or electronic format along with the services designed to help users identify and use those collections Digital libraries promise to provide more effective information services than has been possible in the past, by offering the following advantages: faster delivery, a wider audience, greater availability, more timely information, more comprehensive [8]
- Digital libraries are organised collections of digital content made available to the public. They can consist of material that has been digitised, such as digital copies of books and other 'physical' material from libraries and archives. Alternatively, they can be based on information originally produced in digital format [9]

What is a Digital Library ?



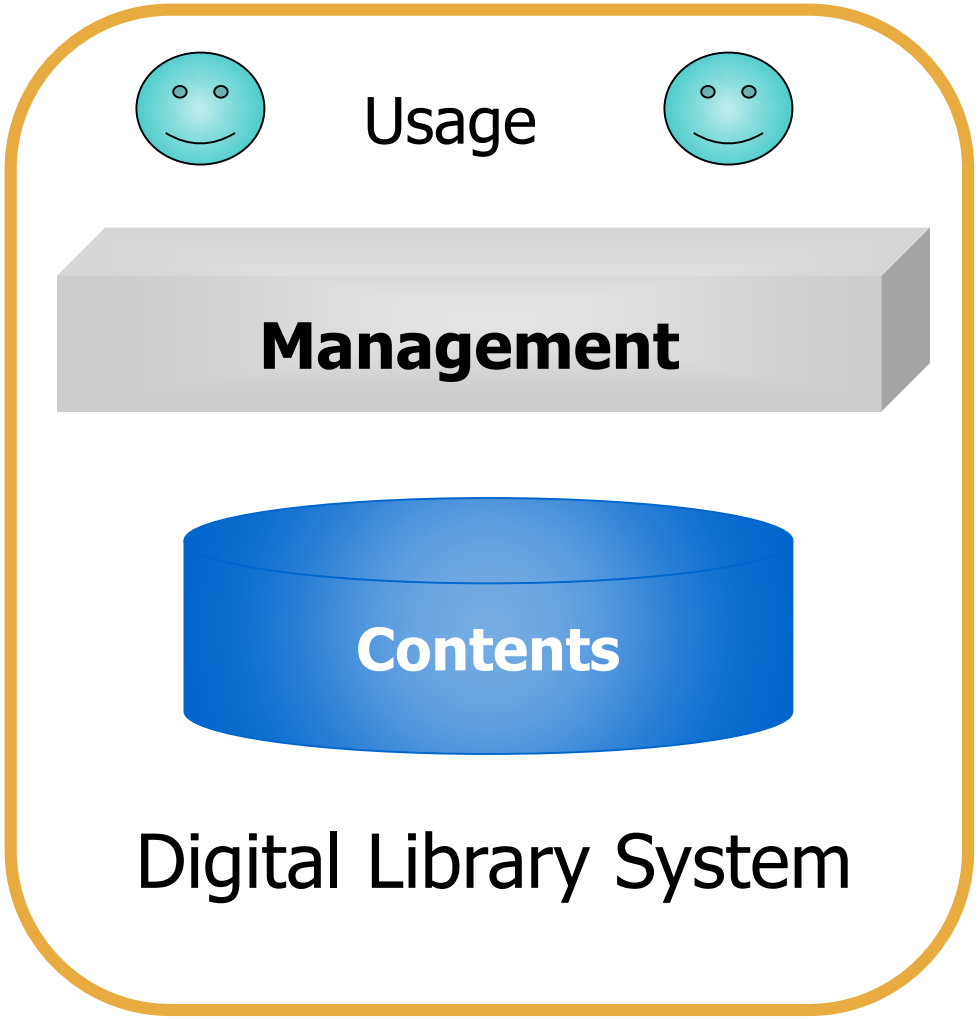
- A DL is the combination of content and services
- A Digital Library is a tool at the centre of intellectual activity having no logical, conceptual, physical, temporal, or personal borders or barriers on information
- A DL is an entity providing the functionality to mediate between information objects and information users in the context of distributed collections of information objects. This (external) functionality includes access, publish, delivery, preservation, personalization, etc.
- A Digital Library is an institution in charge of providing at least the functionality of a traditional library in the context of distributed and networked collections of information objects

Need for a conceptual model

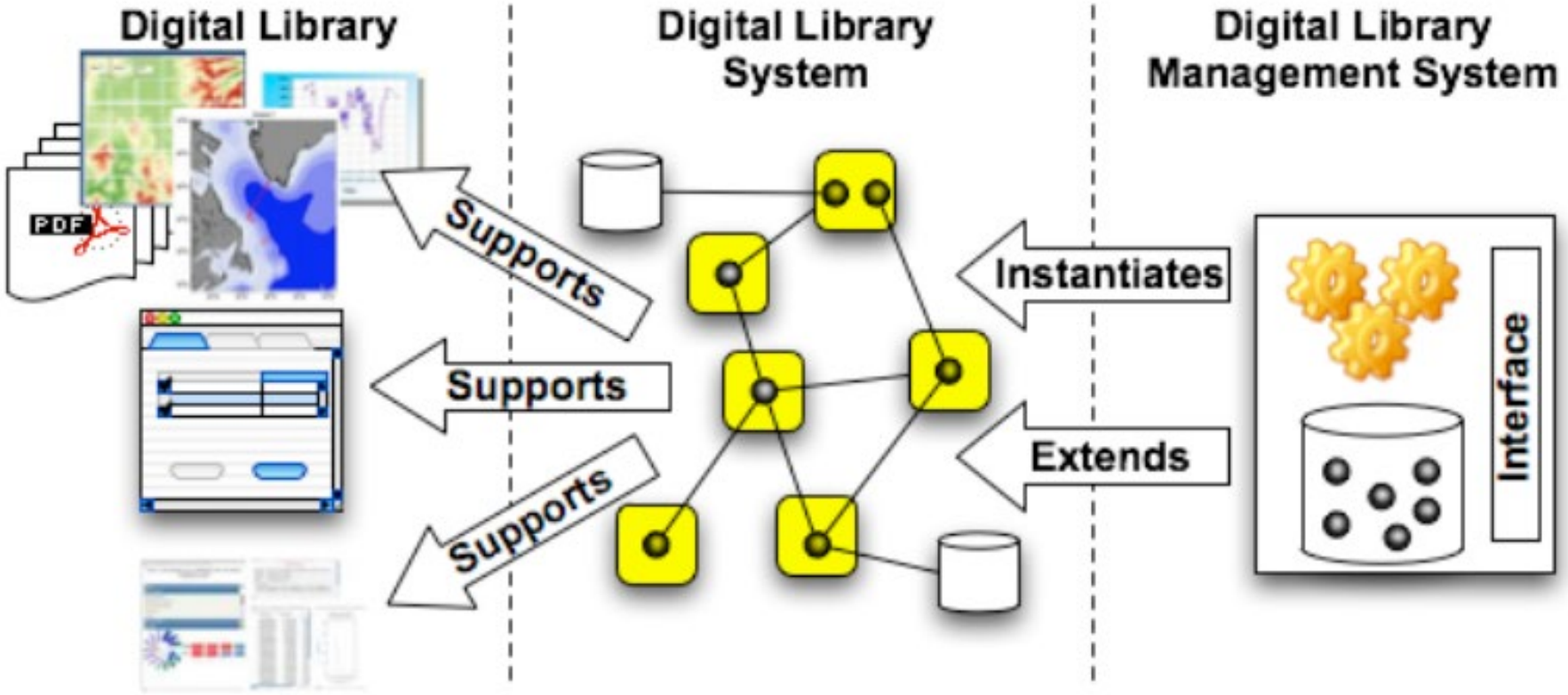


- A conceptual model is an **abstract framework** for understanding significant relationships among the entities of some environment, and for the development of consistent standards or specifications supporting that environment
- A conceptual model is based on **a small number of unifying concepts** and may be used as a basis for education and explaining standards to a non-specialist
- A conceptual 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
- The DELOS Reference Model model for Digital Libraries
  - Entities of interest in the “Universe of Digital Libraries”
  - Basic concepts and actors in Digital Libraries
  - Conceptual maps of relationships in the Digital Libraries Universe

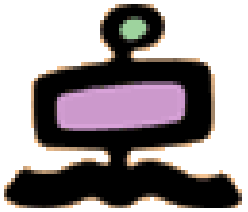
Conceptual Framework



A Three-Entity Framework

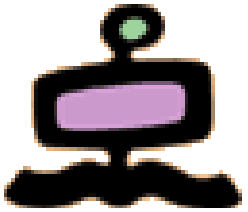


The three Entities



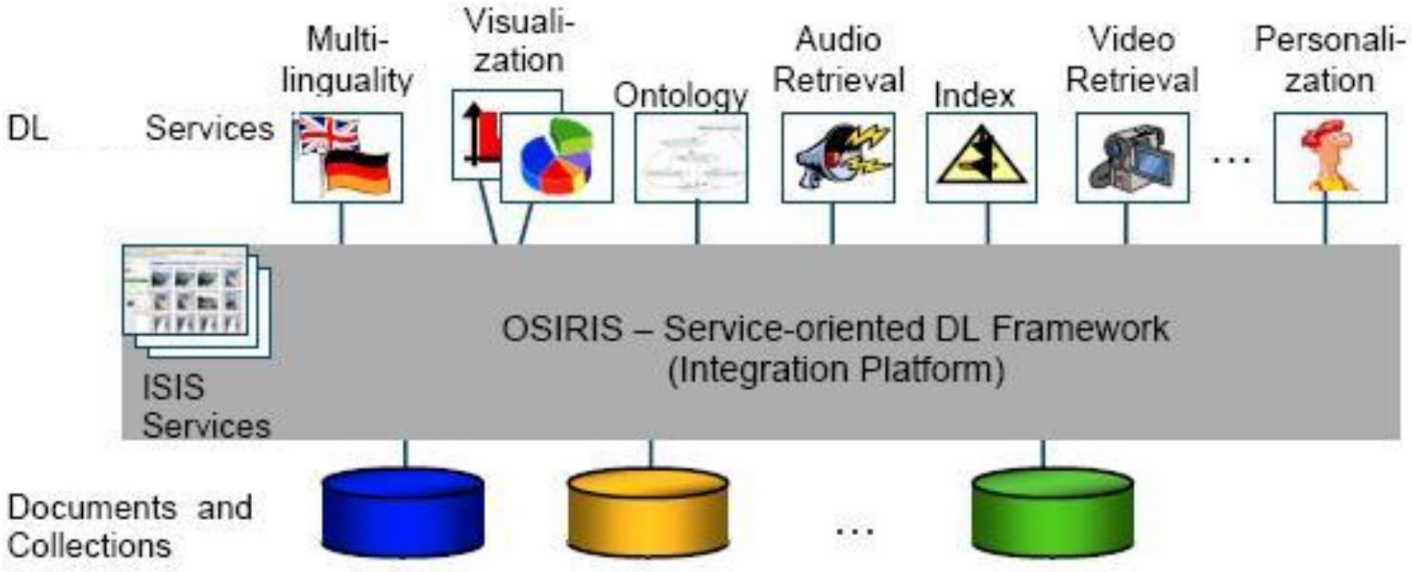
- **Digital Library**
  - An organization, which might be virtual, that comprehensively collects, manages, and preserves for the long term rich **digital content**, and offers to its **user** communities specialized **functionality** on that content, of measurable **quality** and according to codified **policies**
- **Digital Library System**
  - A software system that is based on a defined (possibly distributed) **architecture** and provides all functionality required by a particular Digital Library. Users interact with a Digital Library through the corresponding Digital Library System
- **Digital Library Management System**
  - A generic software system that provides the appropriate software infrastructure both (i) to produce and administer a Digital Library System incorporating the suite of functionality considered foundational for Digital Libraries and (ii) to integrate additional software offering more refined, specialized, or advanced functionality

Different types of DLMS

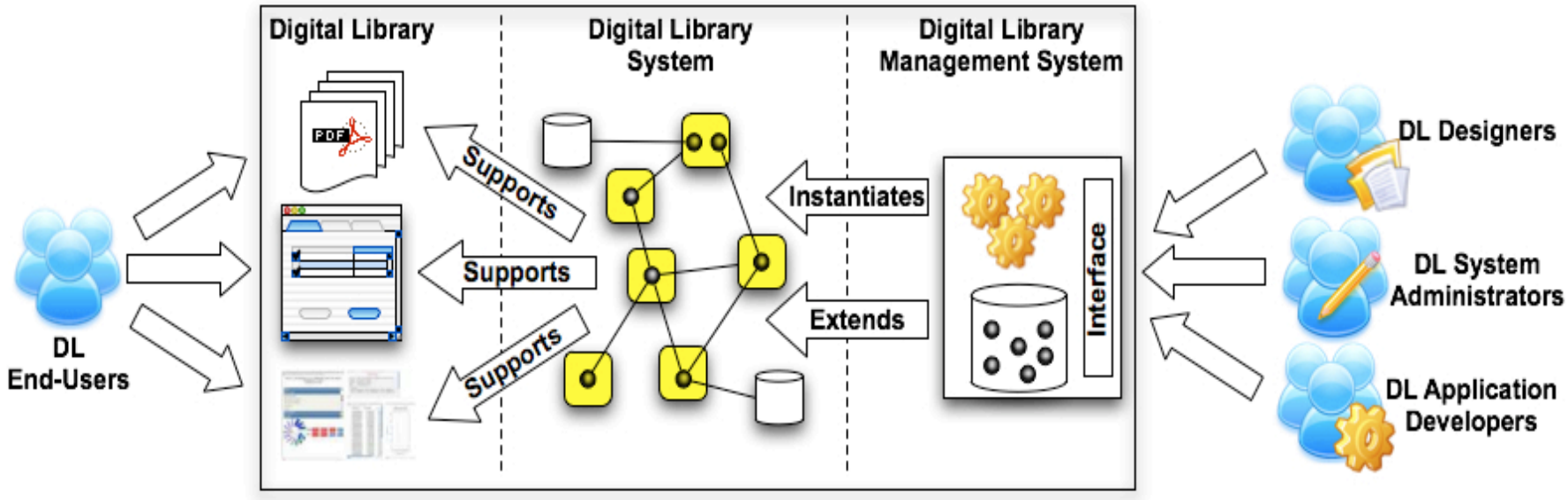


- **Extensible Digital Library System**
  - A complete Digital Library System that is fully operational with respect to basic/ foundational functionality required. It is based on an open software architecture, so that further software components can be incorporated on top of the ones already there with ease (DelosDLMS, Greenstone)
- **Digital Library System Warehouse**
  - A collection of software components that encapsulate the core suite of DL functionality and a set of tools that can be used to combine these components in a variety of ways (in Lego-like fashion) to create Digital Library Systems offering a tailored integration of functionalities. New software components can easily be incorporated into the Warehouse for subsequent combination with those already there (BRICKS, DILIGENT)
- **Digital Library System Generator**
  - A highly parameterized software system that encapsulates templates covering a broad range of functionalities, including a defined core suite of DL functionality as well as any advanced functionality that has been deemed appropriate to meet the needs of the specific application domain Through an initialization session, the appropriate parameters are set and configured; at the end of that session, an application is automatically generated, and this constitutes the Digital Library System ready for installation and deployment (MARIAN)

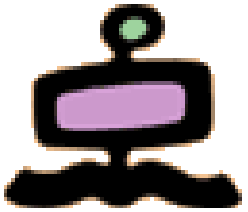
DELOS DLMS


Multi-
linguality
Visuali-
zation
Ontology
Audio
Retrieval
Index
Video
Retrieval
Personali-
zation
DL
Services
...
OSIRIS – Service-oriented DL Framework
(Integration Platform)
ISIS
Services
Documents and
Collections
...

Actors in the Digital Library



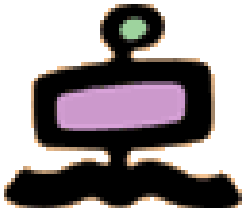
Main roles of Actors (1)



- **DL End-Users**
  - They 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 stateful entity that serves their functional needs The behaviour and output of the DL depend on its state at the time a particular part of its functionality is activated. DL end-users may be further partitioned into
    - Content Creators
    - Content Consumers
    - Librarians (end user)

- **DL Designers (Digital Librarian)**
  - They exploit their knowledge of the semantic of the application 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 The values of these parameters, which can be modified during the DL lifetime, configure the specific DL perceived by the end-users because they determine the particular Digital Library System instance serving the Digital Library

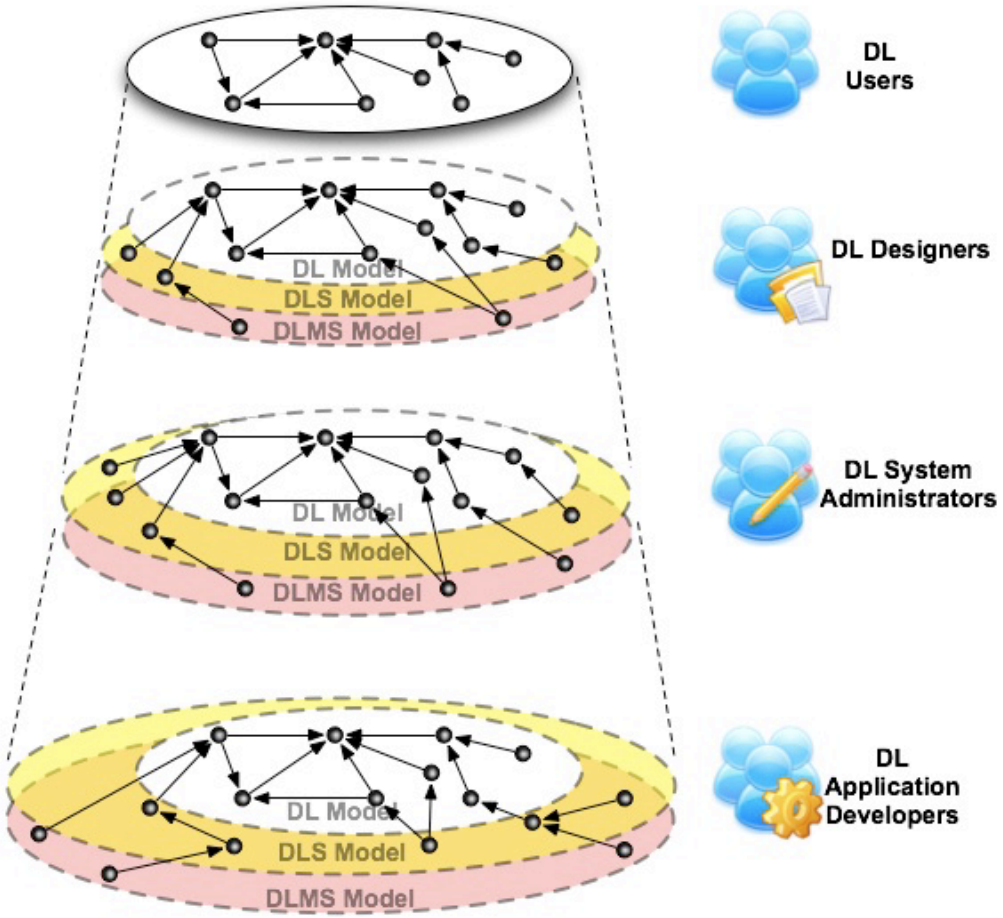
Main roles of Actors (2)



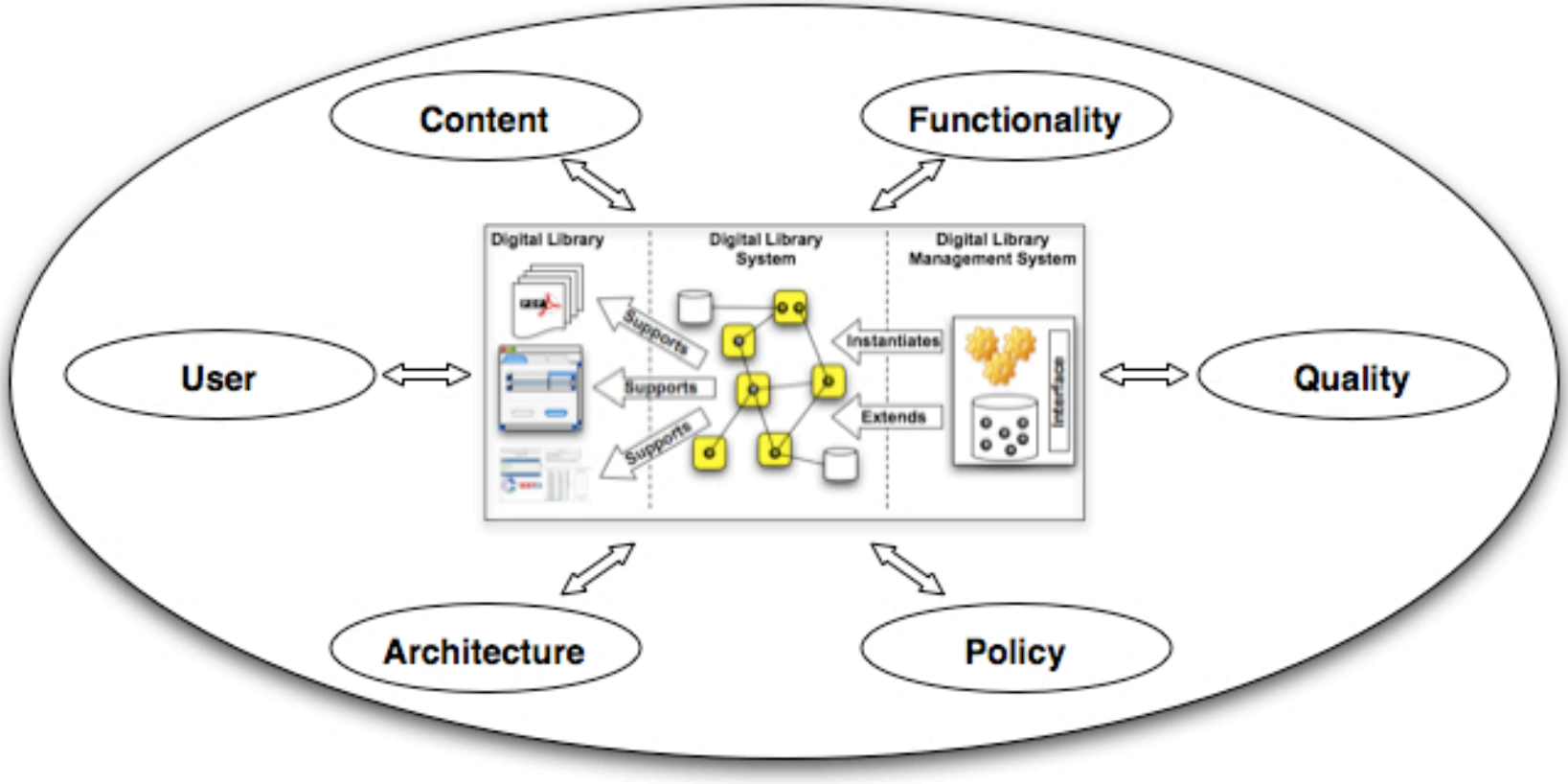
- **DL System Administrators (System Librarian)**
  - They select the software components necessary to create the Digital Library System needed to serve the required DL (as specified by the DL Designer) and decide where and how to deploy them They interact with the DLMS by providing architectural configuration parameters, such as the selected software components, the hosting nodes, and the components allocation. The value of the architectural configuration parameters can be changed over the DL lifetime Any change of these parameters may result in the provision of different DL functionality and/or different quality

- **DL Application Developers**
  - They develop the software components of DLMSs and DLSs, implementing the necessary functionality

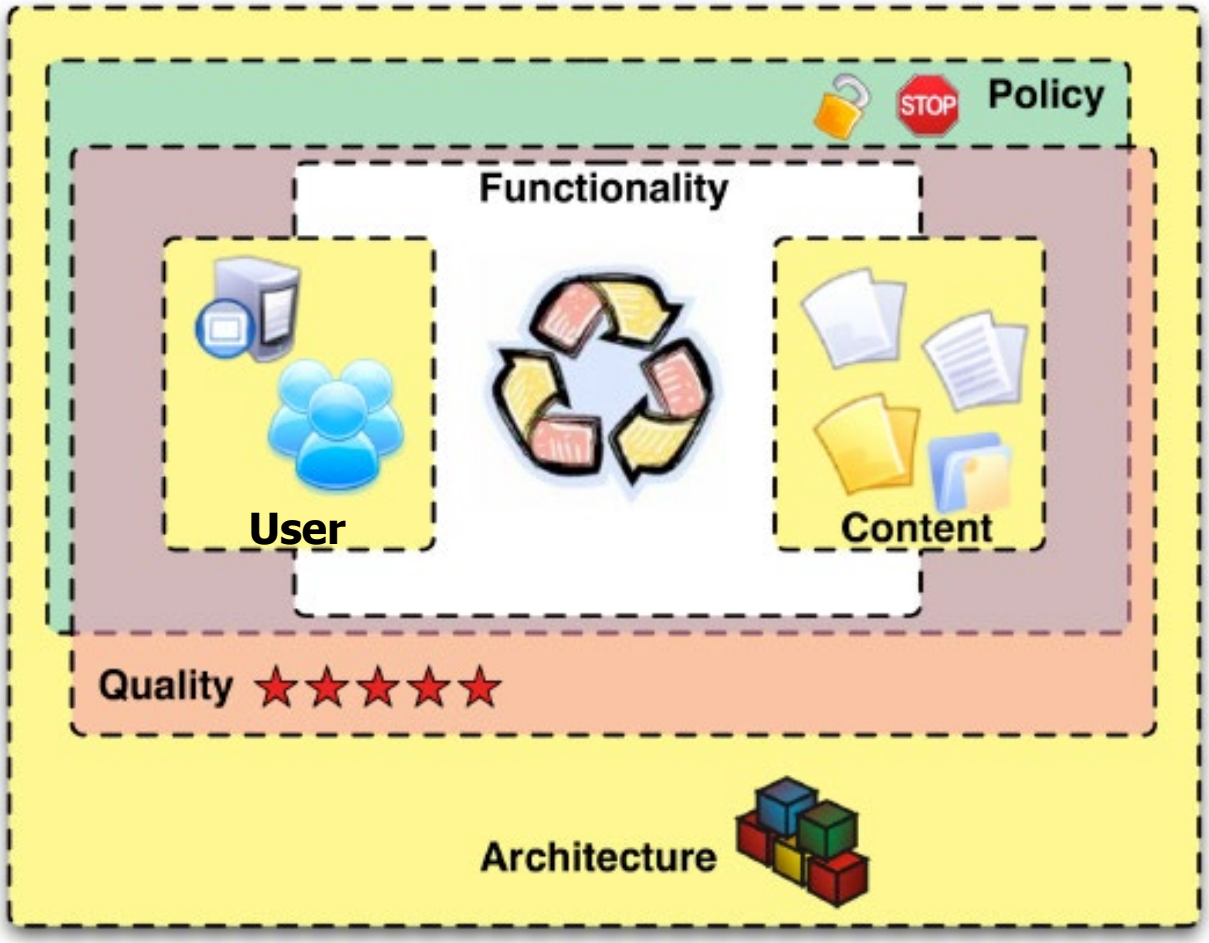
Hierarchy of Actors' Views



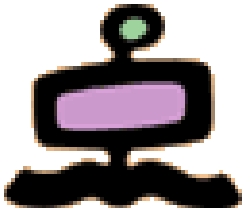
Main concepts (1)



Main concepts in DL

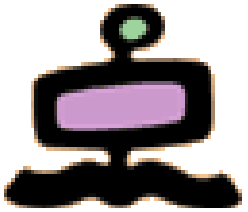


Main concepts (1/3)



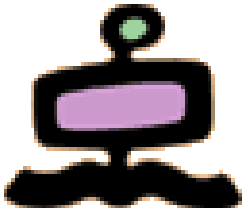
- **Content**
  - The Content concept encompasses the data and information that the Digital Library handles and makes available to its users Content is an umbrella concept used to aggregate all forms of information objects that a Digital Library collects, manages, and delivers It encompasses the diverse range of information objects, including such resources as objects, annotations, and metadata
- **User**
  - The User concept covers the various actors (whether human or machine) entitled to interact with Digital Libraries. Digital Libraries connect actors with information and support them in their ability to consume and make creative use of it to generate new information User is an umbrella concept including all notions related to the representation and management of actor entities within a Digital Library. It encompasses such elements as the rights that actors have within the system and the profiles of the actors with characteristics that personalize the system's behaviour or represent these actors in collaborations

Main concepts (2/3)



- **Functionality**
  - The Functionality concept encapsulates the services that a Digital Library offers to its different users, whether classes of users or individual users While the general expectation is that DLs will be rich in capabilities and services, the bare minimum of functions would include such aspects as new information object registration, search, and browse Beyond that, the system seeks to manage the functions of the Digital Library to ensure that the functions reflect the particular needs of the digital library's community of users and/or the specific requirements relating to the Content it contains
- **Policy**
  - The Policy concept represents the set (or sets) of conditions, rules, terms and regulations governing interaction between the Digital Library and users, whether virtual or real. Examples of policies include acceptable user behaviour, digital rights management, privacy and confidentiality, charges to users, and collection delivery

Main concepts (3/3)



- **Quality**
  - The Quality concept represents the parameters that can be used to characterize and evaluate the content and behaviour of a Digital Library. Quality can be associated not only with each class of content or functionality but also with specific information objects or services Some of these parameters are objective in nature and can be automatically measured, whereas others are subjective in nature and can only be measured through user evaluations
- **Architecture**
  - The Architecture concept refers to the Digital Library System entity and represents a mapping of the functionality and content offered by a Digital Library onto hardware and software components There are two primary reasons for having Architecture as a core concept: (i) Digital Libraries are often assumed to be among the most complex and advanced forms of information systems; and (ii) interoperability across Digital Libraries is recognized as a substantial research challenge A clear architectural framework for the Digital Library System offers ammunition in addressing both these issues effectively