

# Introduction to Digital Libraries and their Technologies



Miroslav Bartošek

Institute of Computer Science MU

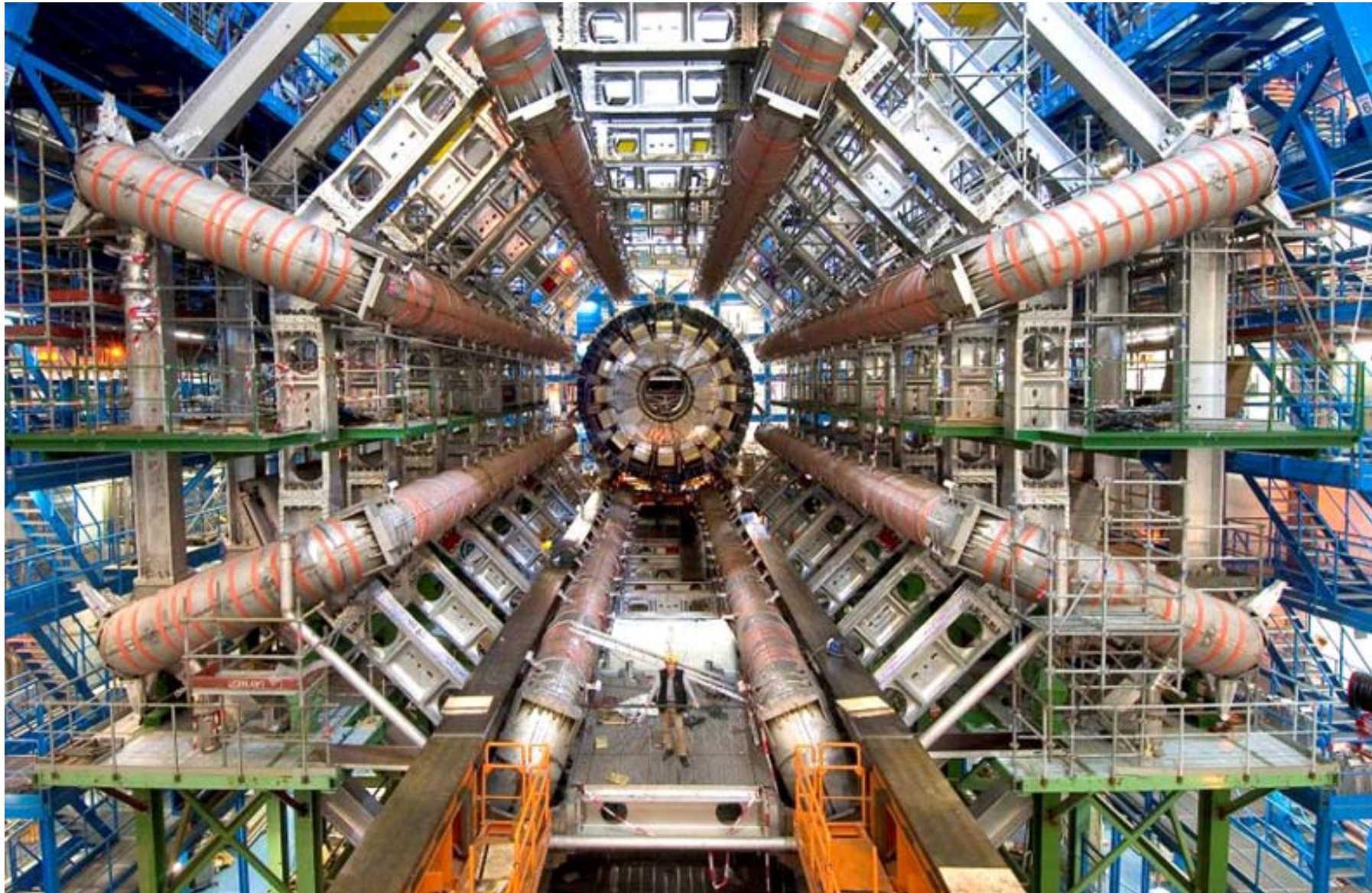
Library and Information Centre

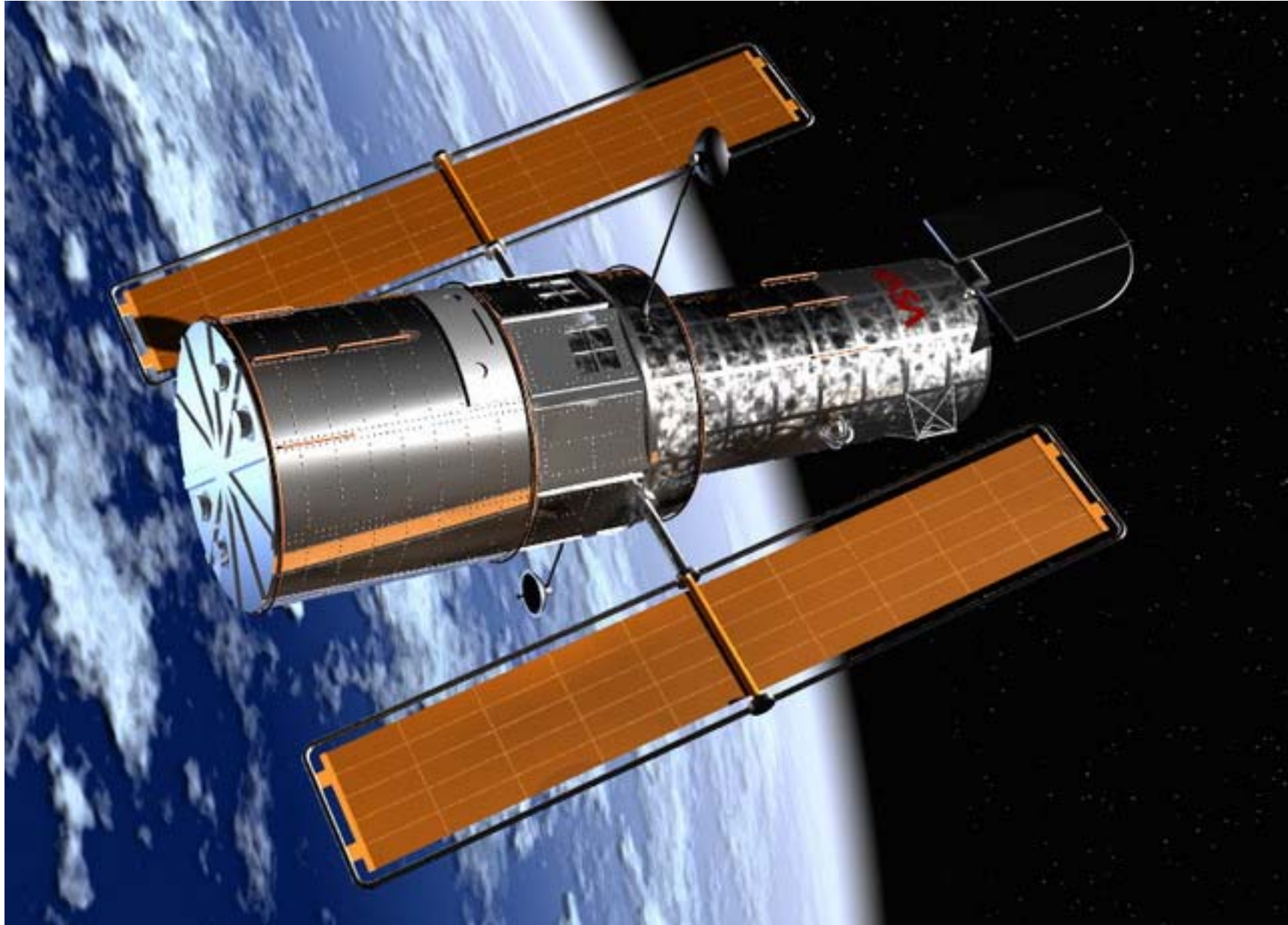
## Picture Quiz: Can you recognise ...











**What do all these system have  
in common?**

**All is about INFORMATION!**

# Topics

1. Introduction to Digital Libraries (DLs)
2. Architecture of DLs
3. Identifiers
4. Metadata
5. Interoperability
6. Searching
7. Economy and Legislation
8. Digital Preservation
9. DLs@MU



# 1. Introduction to Digital Libraries



Logo from [www.ncstrl.org](http://www.ncstrl.org)

# 1. DL Introduction

1.1 DL definition

1.2 DL Examples

1.3 DLs versus WEB

1.4 DLs versus Libraries

1.5 DL's History

1.6 Literature

# 1.1 DL “definition”

## *Computer scientist’s view:*

- Digital library is a **managed collection of information**, with associated **services**, where the information is stored in **digital formats** and accessible over a **network**. *W.Y.Arms, 2000*

- maintained collection
  - services
  - distant access
- 

- **Focused collection of digital objects**, including text, video, and audio, along with methods for **access and retrieval**, and for **selection, organization, and maintenance**. *I.W.Witten, 2002*

- digital content(text, video, audio, 3D, simulation, dynamic visualization)
- user (access and retrieval)
- „librarian“ (selection, organization, and maintenance)

# 1.1 DL “definition”

## *Librarian’s view:*

- 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

*US Digital Library Federation, 1997*

- DL as an „institution“ (a library, for example)
- organization of information and services
- aimed at a defined user community

# 1.1 DL “definition”

## *Archivist’s view:*

- DL = the **infrastructure, policies** and **procedures**, and organisational, political and economic mechanisms necessary to **enable access** to and **proservation** of digital content. *Ross, 2003*
  - DL as a preservation infrastructure / archives

# 1.1 DL “definition”

- *A digital library* is an online collection of digital objects, of assured quality, that are created or collected and managed according to internationally accepted principles for collection development and made accessible in a coherent and sustainable manner, supported by services necessary to allow users to retrieve and exploit the resources.

**IFLA/UNESCO Digital Library Manifesto, 2011**

**International Federation  
of Library Associations  
and Institutions**



# 1.1 DL General Features

- Organization of information is the key
- Not a single closed entity (-> DLs)
- Heterogeneous, dynamic, multimedial information resources
- Interconnection of autonomous units
- Transparent interconnection
- Coherent access regardless of
  - forms
  - formats
  - locations
- Long-term preservation

# 1.2 DL Example (1)



## American Memory

- Digitization „Apollo project“ (pilot 1990-1994)
- 120+ historical collections
- > 10 millions of digital objects
- Books, photos, manuscripts, audio, video, maps

<http://memory.loc.gov/>

## Library of Congress





# 1.2 DL Example (2)



## JSTOR

- DL of Academic Journals (founded 1995)
- Problems with printed journals in libraries (cost, space, incompleteness, preservat.)
- Idea: digitize all our core journals from 1<sup>st</sup> issue to ... **moving wall**
- Non-profit organization, Mellon F grant (Ann Arbor, New York)
- 1900 digitized journals, 900 publishers
- 20 subject collections (arts, sci, soc-sci)
- Sustainable economic model
- 8.000+ institutions from 160 countries

<http://www.jstor.org/>

## Journal Storage

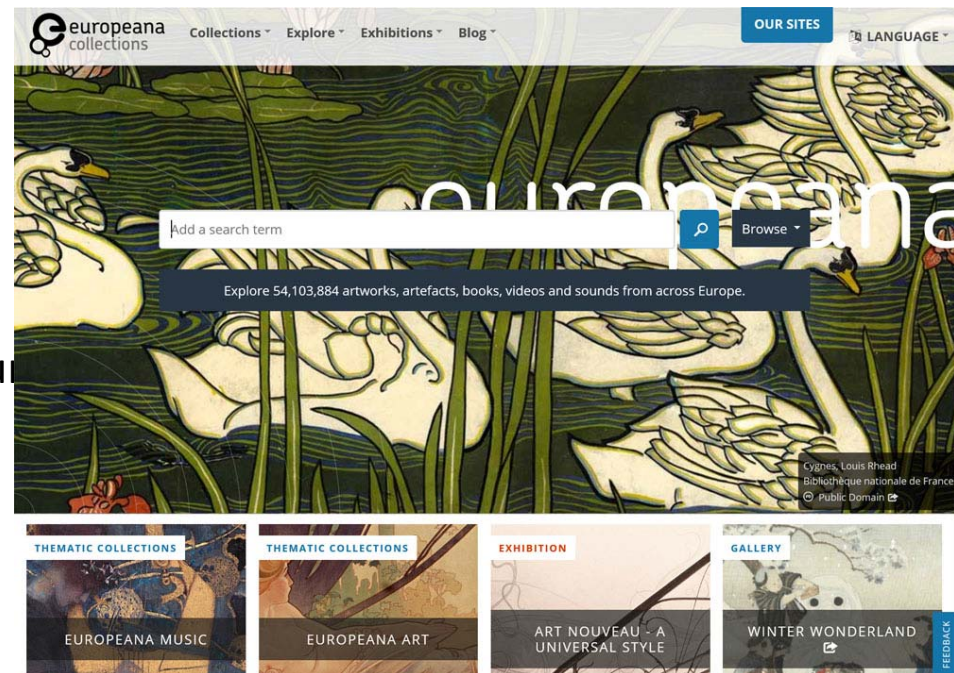
A screenshot of the JSTOR website homepage. The header is dark red with the JSTOR logo on the left and navigation links for 'FOR LIBRARIANS', 'FOR PUBLISHERS', and 'FOR INDIVIDUALS' on the right. Below the header, there's a tagline 'Light up your mind. Scholarly journals, primary sources, and now books!'. The main content area features three vertical panels: 'FOR PUBLISHERS' with a stack of books, 'FOR INDIVIDUALS' with a portrait of a man, and 'FOR LIBRARIANS' with a library aisle. To the right of these panels is a red box with the text 'The seeds of something new' and a search bar. Below this, there are sections for 'THE BUZZ' with social media links, 'JSTOR EVIDENCE IN UNITED STATES VS. SWARTZ' with a 'Read more' link, and '10 THINGS YOU SHOULD KNOW ABOUT JSTOR' with a list of two points. On the far right, a 'JSTOR IN NUMBERS' box displays '151M Searches Performed in 2012'.

# 1.2 DL Example (3)



## Europeana

- EU digital platform for cultural heritage
- Initiated and supported by EC
- 2010 first version
- 2016:
- 3.500 contributing institutions  
Libraries, museums, archives, galleries...
- 54 mil of objects
  - 30 pictures, 22 texts, audio, video
- Metadata, thumbnail + link to resource



<http://europeana.eu/>

## 1.3 DL x Web

- ***Why DL?***  
We have the Web and there is „all“ !
- Is the Web a digital library ?

# 1.3 The web is great, but...

- Huge amount of information, easy access to anybody
- Unified technology
- Continuous exciting development
- And much more...

## But:

- Advanced and Non-textual search
- Rights Management
- Permanent Availability
- Authenticity
- Quality Control

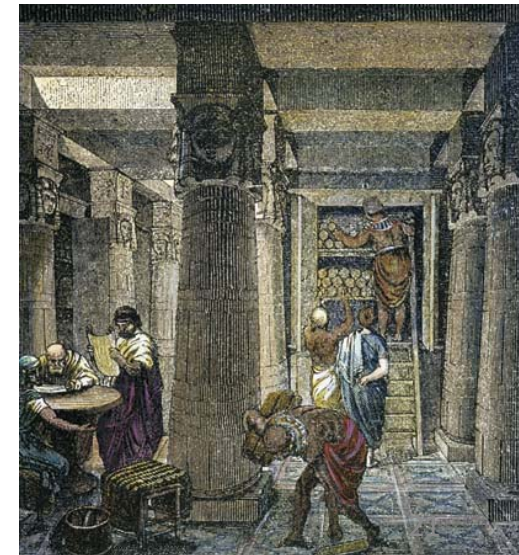


# 1.4 DLs x Libraries

## Common features

- Systematically built collection of data objects
- Metadata structures (catalogues, indexes)
- Services tailored to designated user community
- Thematic focus
- Quality Control
- Long-Term Storage (centuries – in libraries)

**Library of Alexandria:** Egypt - Ptolemaics, 295 BC -- ??  
700.000 papyrus scrolls (originals from Euripides, Aeschylus,  
Sophocles, Archimedes, Euclid...)



# 1.4 Advantages of libraries

- Centuries old tradition in organization and access to info
- Worldwide standards
- Elaborated system of libraries
- Established legislation
- Well-balanced system of all key players  
authors – publishers – libraries – users

# 1.4 Transformation to DLs

- „Paper libraries would disappear by 1984.“

Arthur Samuel (1964, The Banishment of the Paperwork.)

- „Some say that had books been invented after computers were, they would have been hailed as a great advance.“

Ian H.Witten (2002, How to Build a Digital Library.)

- **Transformation to DLs**
  - It's not just the technological issue
  - Human beings & social environment are main obstacles

# 1.5 DL's History

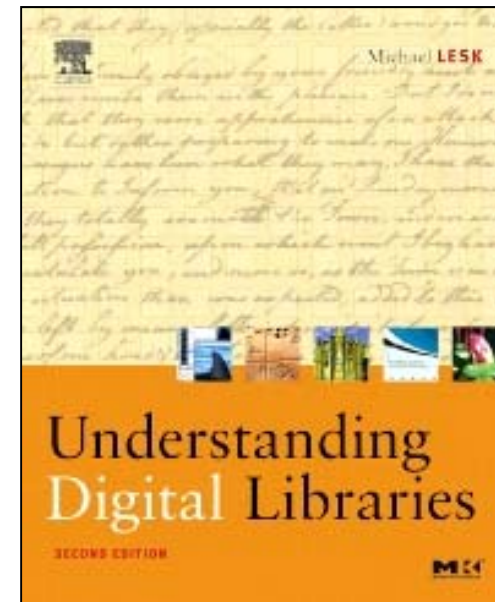
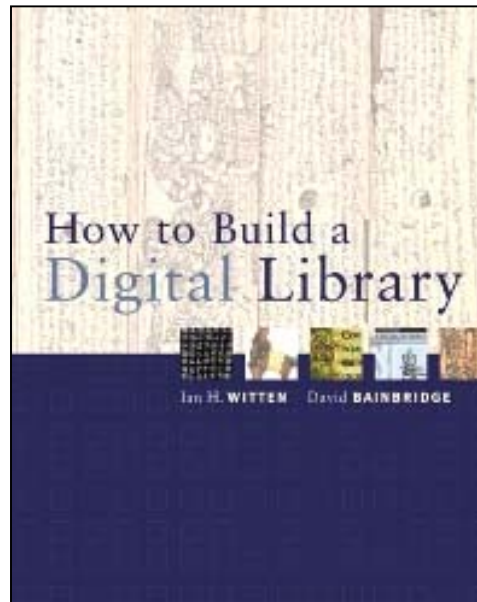
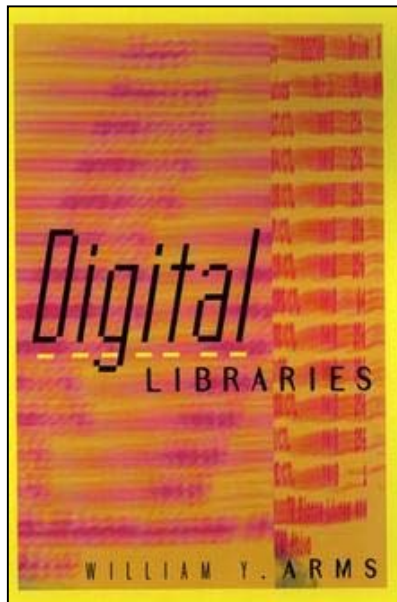


- 1945 Vannevar Bush („As We May Think“, Memex)
  - 1965 J.C.R. Licklider („Libraries of the Future“)
  - 60's MARC, OPAC (LoC, OCLC)
  - 80 's fulltexts
- 
- 90 's **Computing + Communications + Contents**  
(price, performance, availability)
  - 1994 Digital Library Initiative , www



# 1.6 Literature

- William. Y. Arms: **Digital Libraries**. MIT Press, 1999, 2000, 2001  
Online edition (2005) available at <http://www.cs.cornell.edu/wya/diglib/>
- I.H.Witten: **How to Build a Digital Library**. Morgan Kaufmann Publ. 2002, 2010
- Michael Lesk: **Understanding Digital Libraries**. Morgan Kaufmann, 2<sup>nd</sup> ed. 2004



## 2. Architecture of DL



## 2. Architecture of DLs

2.1 Reference Models

2.2 Kahn-Wilensky Framework

2.3 DL.org Model

# 2.1 Reference Models

**Reference Model** = general architecture (framework)

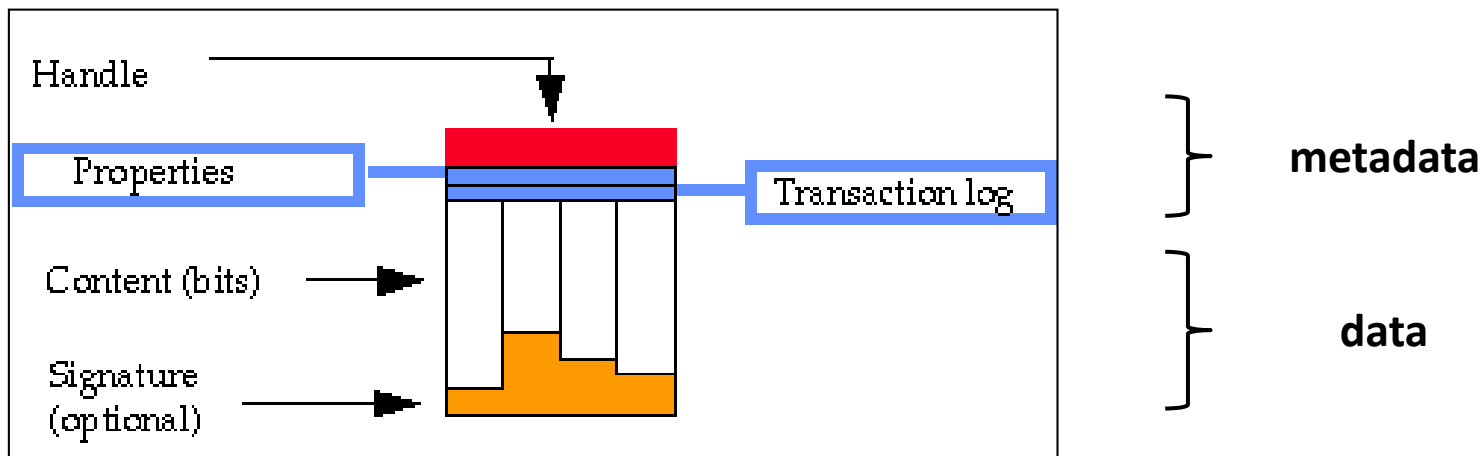
- Provision of a unified vocabulary (terms)
- Formalizing components and functions (semantics)
- Understanding important relationships between entities in a particular environment
  - Software implementation
  - Development of standards
  - Education



# 2.1 Kahn-Wilensky Framework

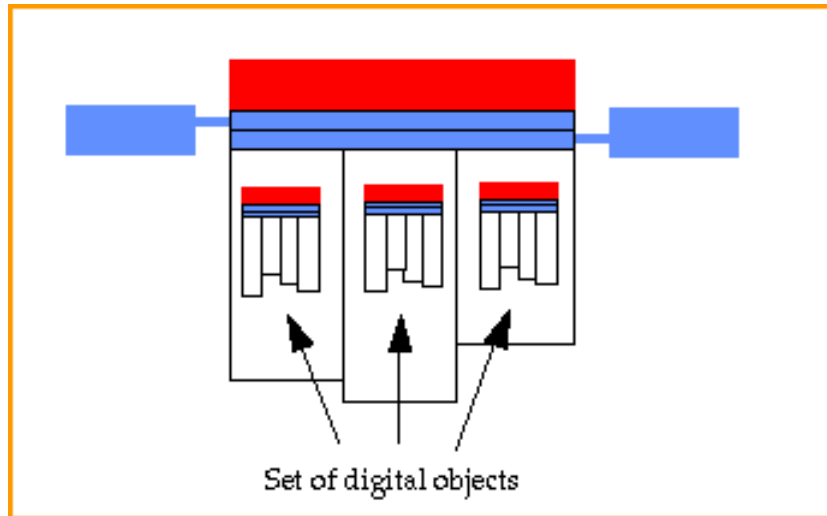
- First informal model/architecture for the DL
- R.Kahn, R.Wilenski: [A Framework for Distributed Digital Object Services](#), Uni Berkeley, CS-TR project, ARPA, 1995  
<http://www.cnri.reston.va.us/home/cstr/arch.html>
- Digital object; Identification system (handles); Repository; Services
- Implemented in FEDORA

# 2.1.1 Digital Object



## 2.1.2 Composed DO, Meta DO

### Composed DO



Example: **DO=book**

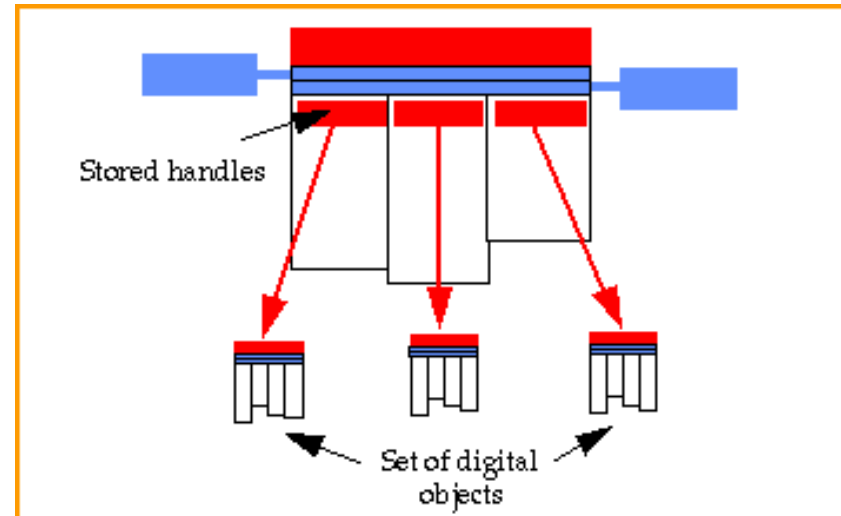
metadata

- ...

data

- DO for page1
- DO for page2
- ...

### Meta DO



Example: **DO=music composition**

metadata

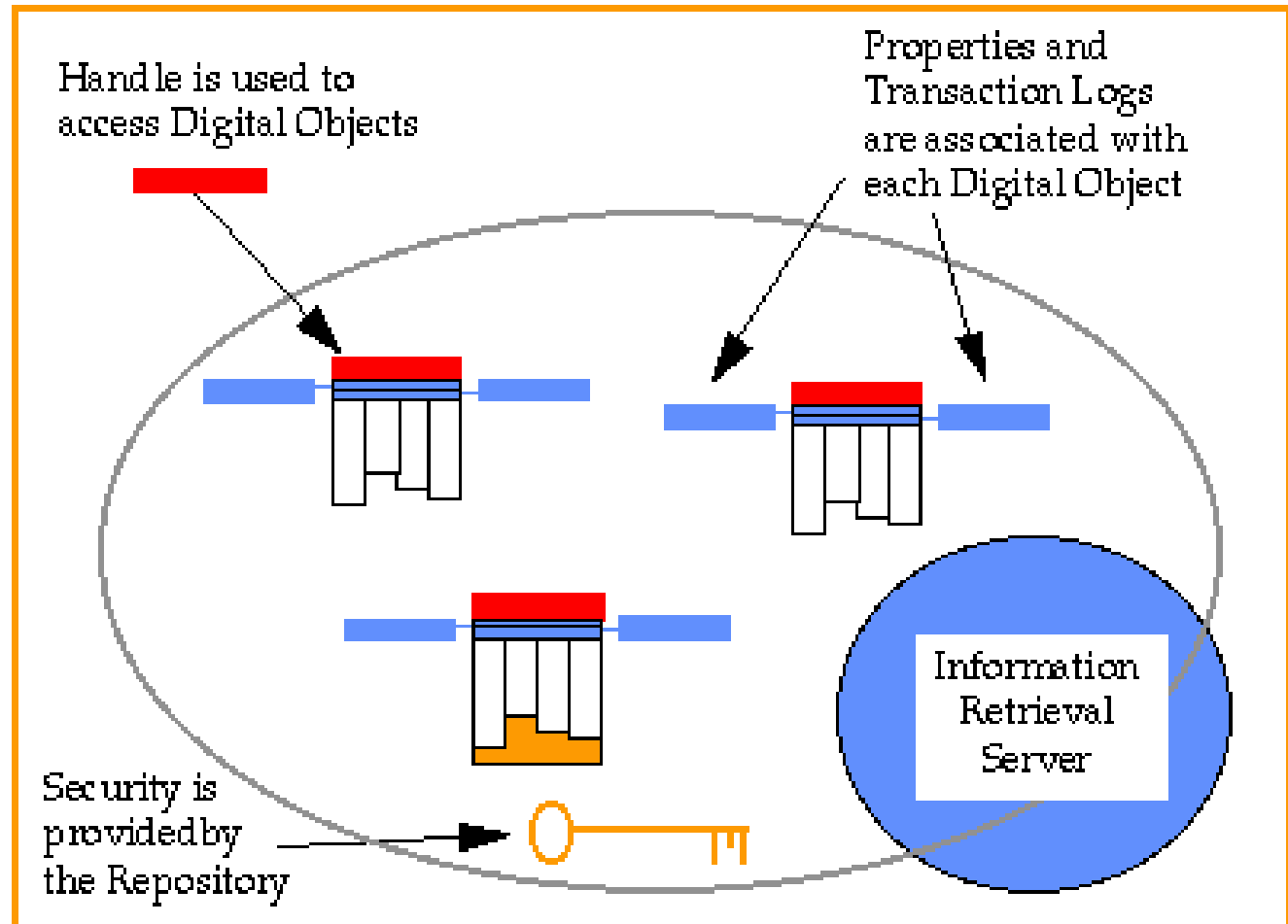
- ...

data

- id for DO score
- id for DO audio recording
- id for DO performance TV record

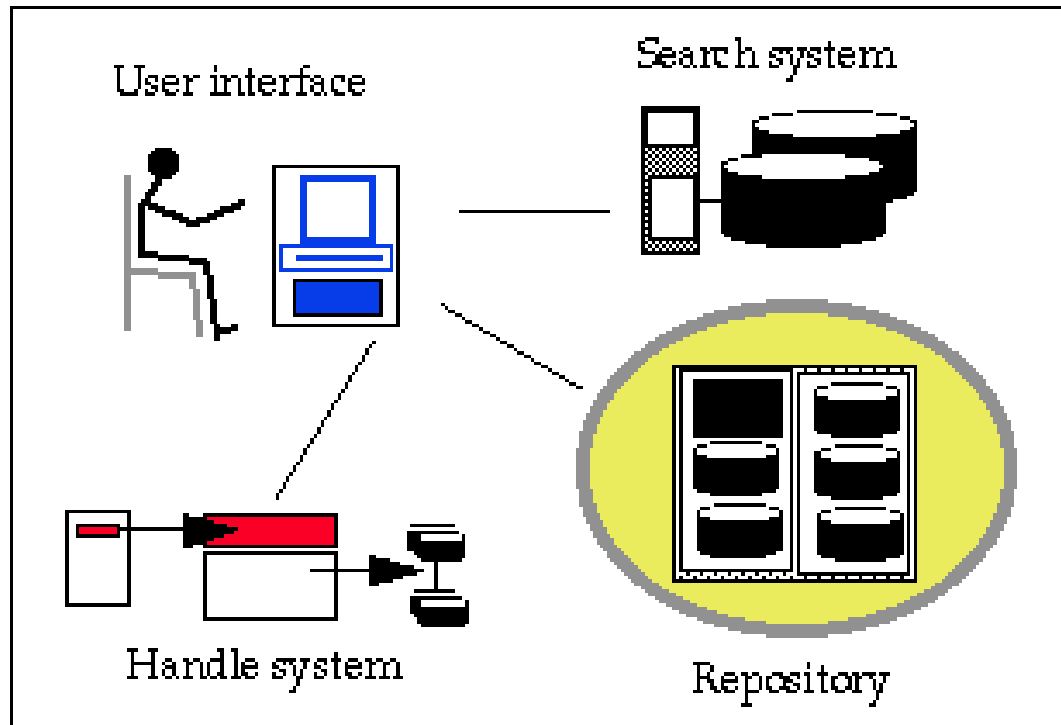
## 2.1.3 Repository

- Takes care of the DOs stored in it
- **RAP**  
Repository  
Access  
Protocol





## 2.1.4 DL Components



- |                    |                |                       |                   |
|--------------------|----------------|-----------------------|-------------------|
| 1. <b>search</b>   | User interface | => Search system      | => list of items  |
| 2. <b>select</b>   | User interface | => Item               | => handle         |
| 3. <b>retrieve</b> | User interface | => Handle system      | => repository ID  |
|                    | User interface | => Repository - RAP   | => digital object |
| 4. <b>display</b>  | User interface | => <i>rendered DO</i> |                   |

## 2.2 DL.org Model

Formal DL model (2011). Result of several EU-funded projects:

- **DELOS**

Network of Excellence on Digital Libraries 2004-2008

<http://www.delos.info>

- **DL.org**

Coordination Action projekt EC 2008-2011

Digital Library Interoperability, Best Practices & Modelling  
Foundations

<http://www.dlorg.eu>

## 2.2 Documentation

### Set of documents

- Digital Library Manifesto
- Digital Library Reference Model
- Digital Library Technology and Methodology Cookbook
- Digital Library Conformance Checklist



### **Digital Library Reference Model** (very extensive document)

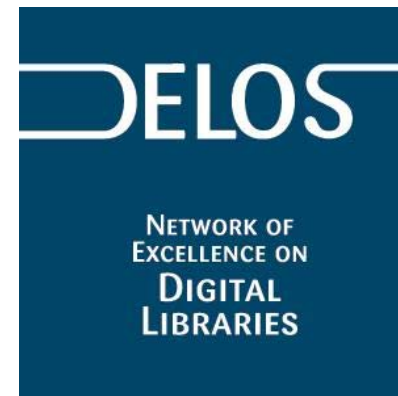
<http://www.dlorg.eu/index.php/outcomes/reference-model>

### Concise booklets:

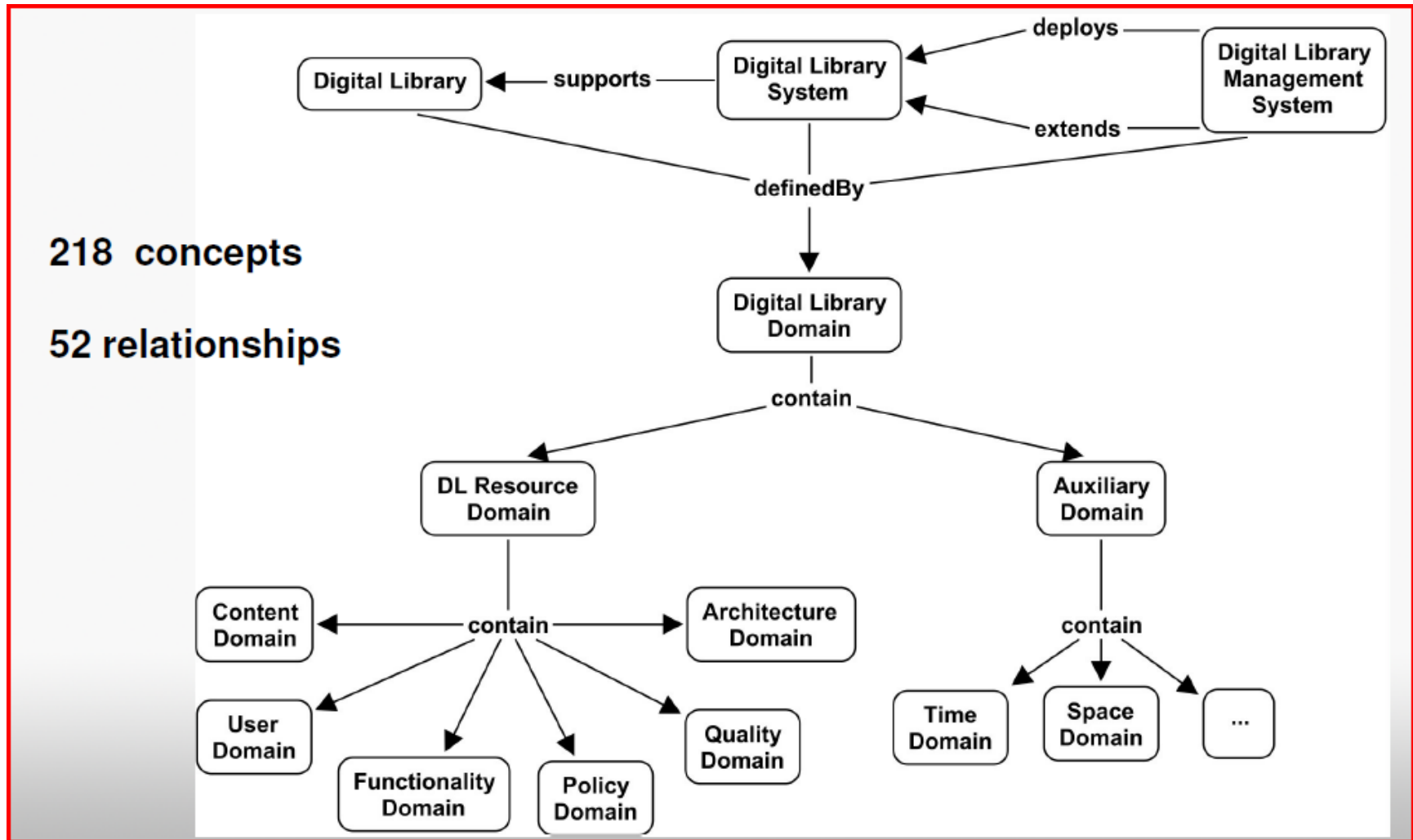
- **DL Reference Model in Nutshell** (16 pages only)  
[http://www.dlorg.eu/uploads/Booklets/booklet21x21\\_nutshell\\_web.pdf](http://www.dlorg.eu/uploads/Booklets/booklet21x21_nutshell_web.pdf)
- **Digital Library Manifesto**  
[http://www.dlorg.eu/uploads/Booklets/booklet21x21\\_manifesto\\_web.pdf](http://www.dlorg.eu/uploads/Booklets/booklet21x21_manifesto_web.pdf)
- **Digital Library Cookbook**  
[http://www.dlorg.eu/uploads/Booklets/booklet21x21\\_cookbook.pdf](http://www.dlorg.eu/uploads/Booklets/booklet21x21_cookbook.pdf)

## 2.2 DELOS – DL vision

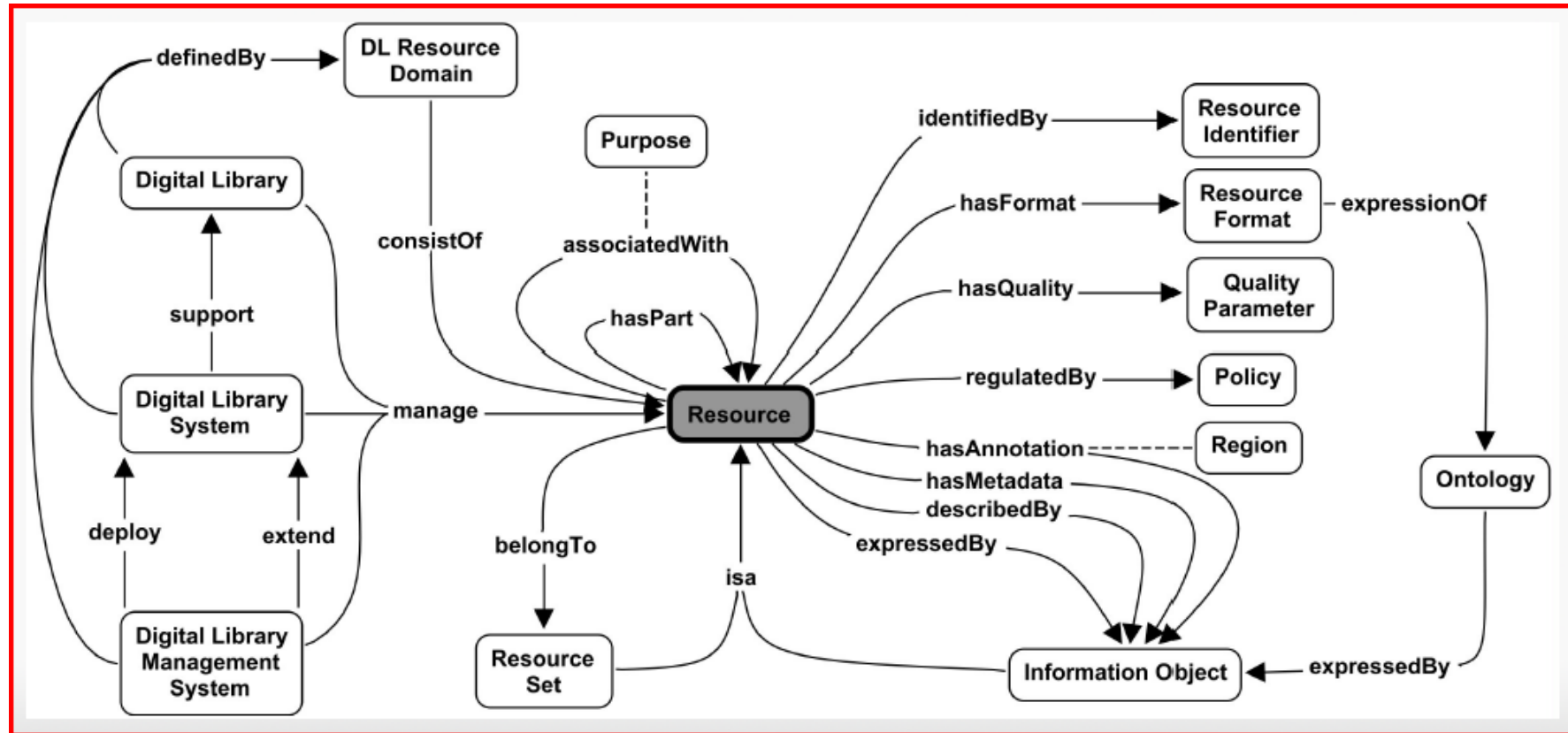
- Digital libraries should enable **any citizen** to access **all** human knowledge **anytime** and **anywhere**, in a **friendly, multi-modal, efficient, and effective** way, by overcoming barriers of **distance, language, and culture** and by using multiple **Internet-connected** devices



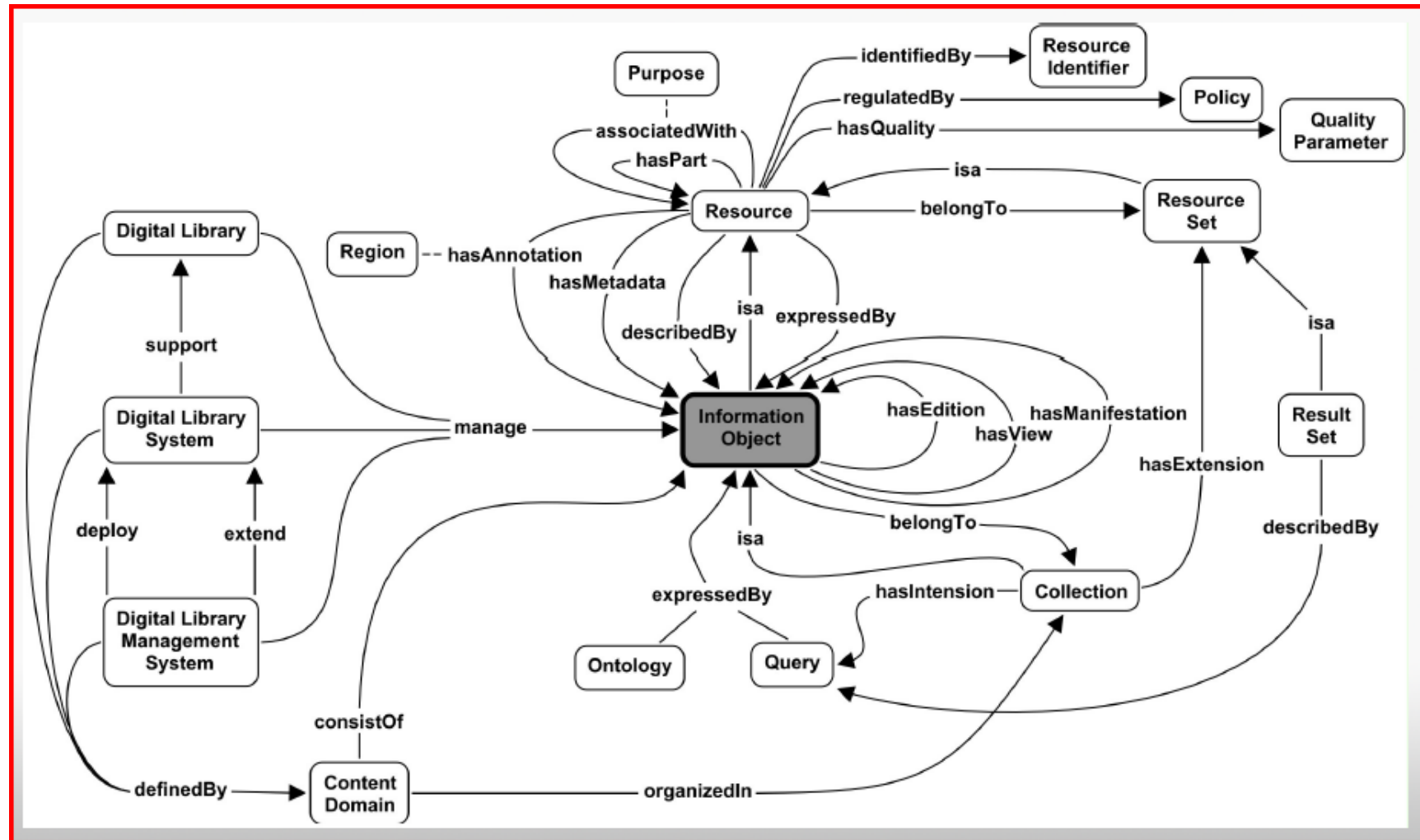
# 2.2 DL Domains



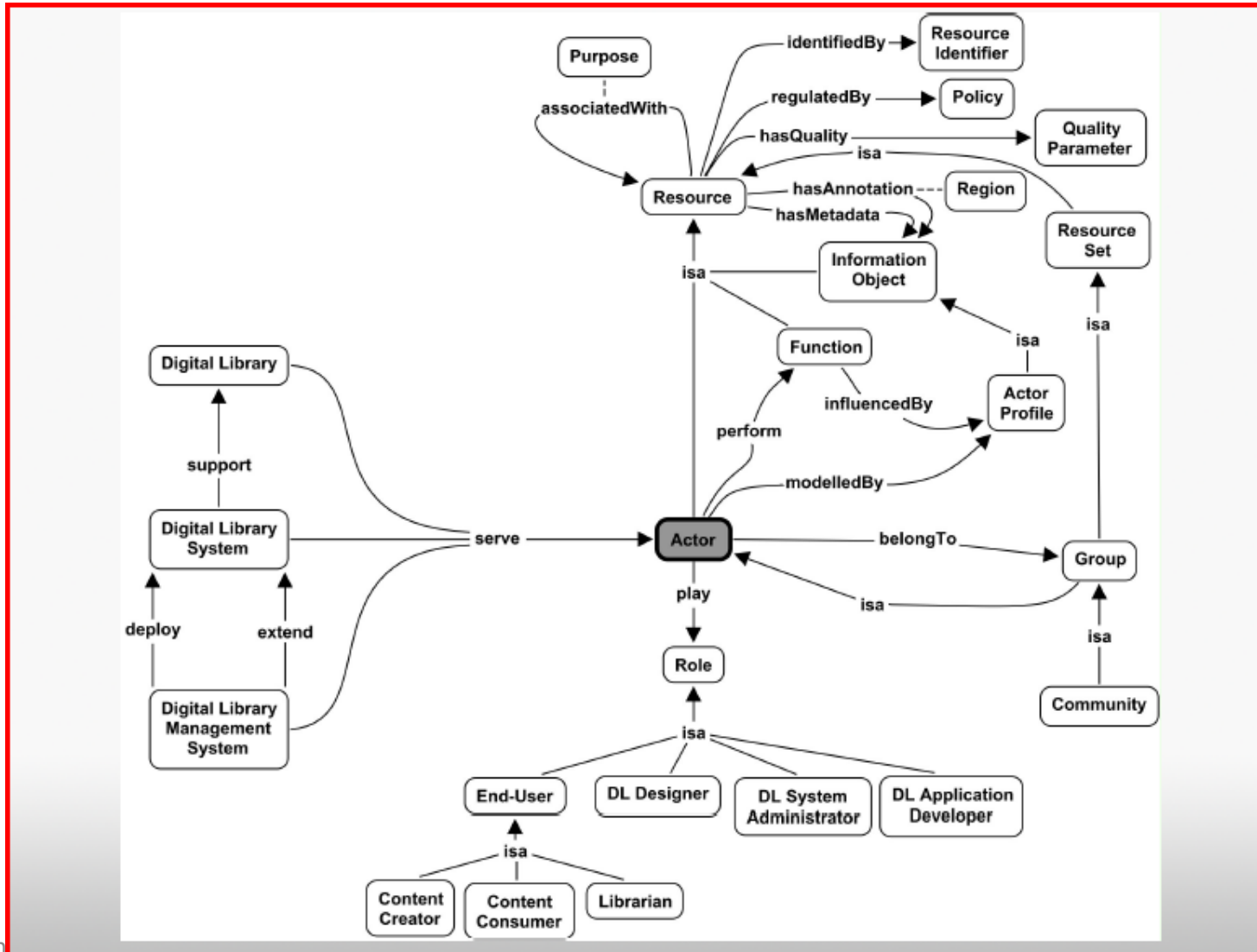
# Example: Resource



# Example: Information Object

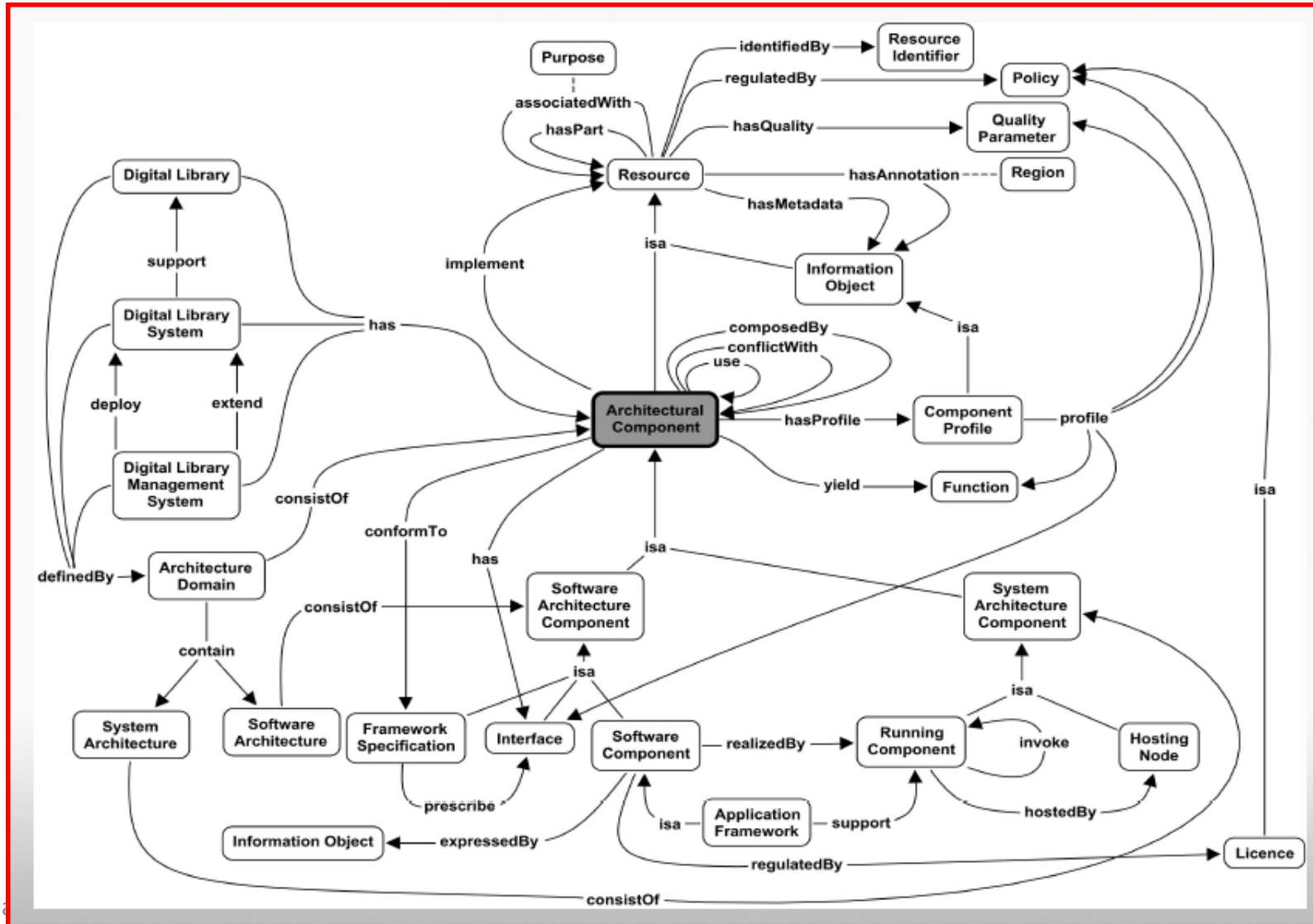


# Example: User

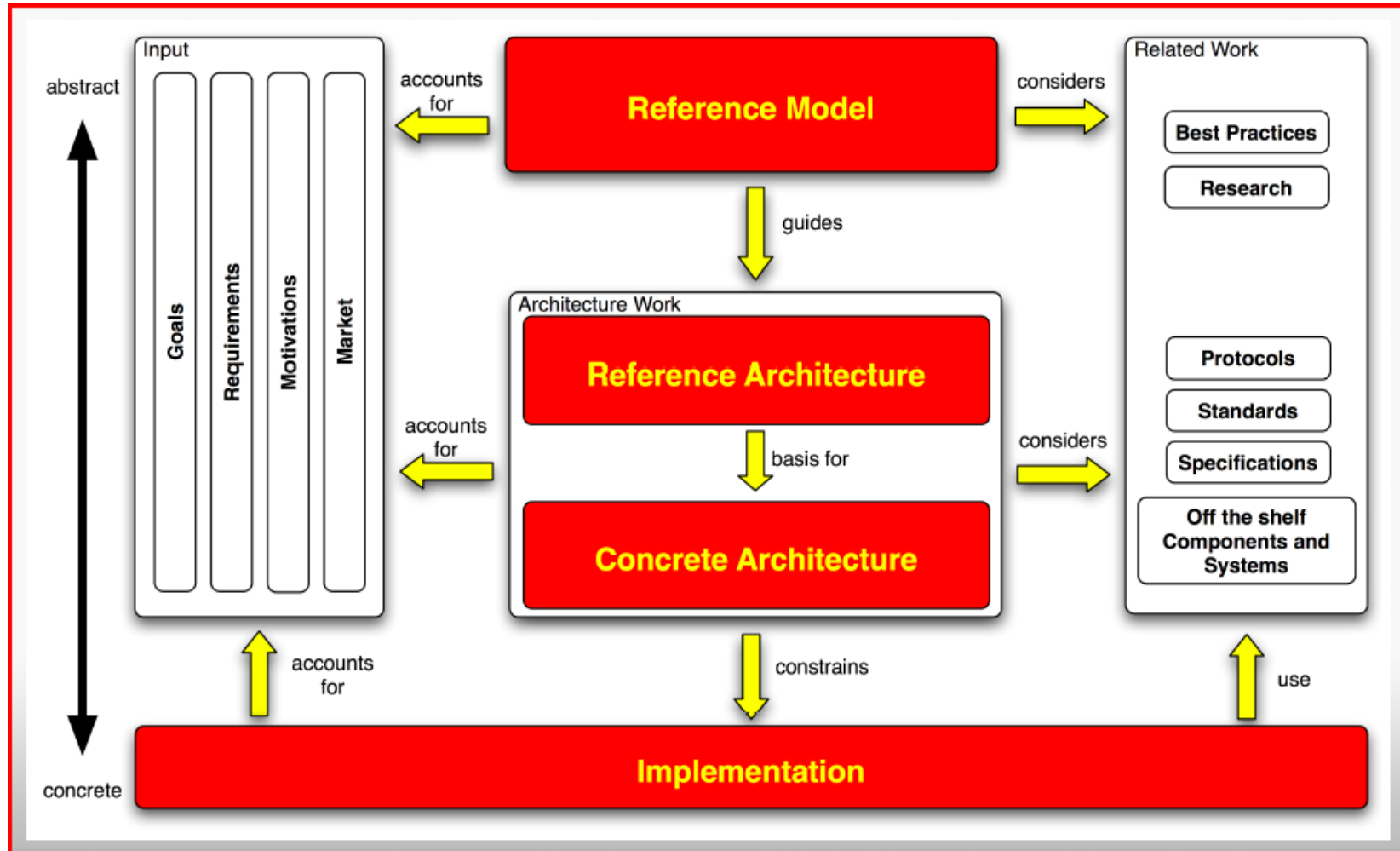




# Example: Architecture



# Design and implementation of DL



# 3. Identifiers



# 3. Identifiers

## 3.1 Introduction, properties of IDs

## 3.2 Classic Library IDs

ISBN, ISSN, SICI, ..., ISTC, ISNI, ...

## 3.3 Digital IDs

URN, PURL, Handles, DOI, ARK

# 3.1 Identifiers

- If there is one thing that distinguishes a digital library from a mere web site, it is that libraries do their best to provide *reliable, persistent access through durable links*.  
(J.A.Kunze, California Digital Library)

- **Identifiers**

- Unique names
- Basic building blocks keeping/binding things together

- Local x Global identifiers

- Eliminating physical contact = higher need for identification
  - Precision
  - Reliability
  - (machine) Linking

# 3.1 Properties of identifiers

- 1. Form**  
(structured, dumb, computable)
- 2. Uniqueness (global)**  
(central / distributed assignement)
- 3. Persistency**  
(future validity and interoperability)
- 4. Resolution (action)**  
(machine system providing for ID its DO, clicable)

# 3.1 Hierarchical system of IDs

No one universal ID for everything => hierarchical system

- **Organisations (library)**

  - ISIL** International Standard Identifier for Libraries and Related Organizations

- **Collection, service**

  - ISCI** Intl Standard Collection Identifier

- **Author**

  - ISNI** Intl Standard Name Identifier

- **Work**

  - ISTC** Intl Standard Text Code

  - ISWC** Intl Standard Musical Work Code

- **Manifestation of work**

  - ISBN** Intl Standard Book Number

  - ISSN** Intl Standard Serial Number

  - ISMN** Intl Standard Music Number

- **Component/article**

  - SICI** Serial Item and Contribution Identifier

  - DOI** Digital Object Identifier

Interantional standards  
(mostly ISO)

## 3.2 Examples: ISBN

- **International Standard Book Number**      **ISBN 80-00-01987-6**
- Classic library identifier, ISO standard since 1972
- Structured id, fixed length, distributed assignment
- Invented for printed environment – very successful, heavily used, useful (publishers, business, libraries, citations)
- **BUT: Serious problems in digital environment**
  - Web publishing – rapid increase of id-requirements
  - Exhaustion of available number space!
  - ISBN-13: Temporary remedy (rapid fix) -- ISBN 978-80-00-01987-3
  - New ISBN desperately needed
    - It takes a long time to agree on a new global standard
    - It will be very costly to implement it



## 3.3 Examples: SICI

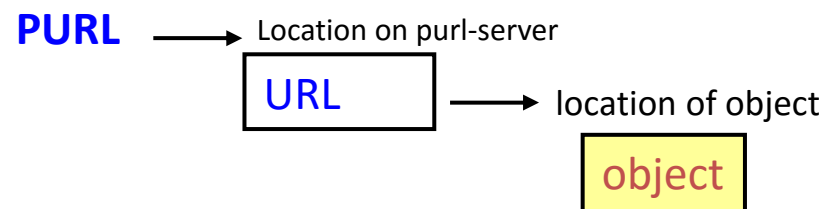
- **Serial Item Component Identifier** – components of journal issue
- [0730-9295\(199206\)11:2<168:CRFAOC>2.0.TX;2-#](#)  
M.Needleman. Computing Resources for an on-line catalog – 10 years later.  
Information technology and libraries. 11(2), červen 1992, pp. 168-175
- Computable id, interesting novel approach
- US ANSI standard since 1996
- **BUT: Didn't gain global acceptance**
  - Replaced by more successful rival – DOI

## 3.4 Filling gaps – new IDs

- **ISNI – International Standard Name Identifier**
  - Global unique identifier for authors, ISO standard since 2012; 9,5 mil assigned
  - **ISNI 0000 0000 7988 7687** (Bartošek, Miroslav)
  - RA: Registration Authority (ISNI International Agency)
  - RAGs: Registration Agencies (currently 12 – British Library, Bibliotheque Nationale de France, ...)
  - ISNI metadata set
- **ISTC – International Standard Textual Code**
  - Intellectual works/creations (expressed mainly in textual form)
  - **ISTC 0A9-2002-12B4A105-6**
  - RA: International ISTC Agency (2008)
  - RAGs: currently 8; 0,2 mil ISTC-ids assigned
  - Huge and costly task to identify all works worldwide – who will do that?

## 3.5 Digital IDs – URL, URN, PURL

- URL – most frequently used on the Web as a “identifier”, BUT:
  - Uniform Resource Locator – identifies location, not an object!
  - Not persistent (broken links – 404 Not Found)
- URN – conceptually known but not deployed
- **PURL** – Persistent URL
  - Pragmatical solution how to improve persistency of URL (OCLC)
    1. PURL is URL
    2. PURL refers to the location where the second URL is located, which refers to the location where the object resides



## 3.6 Handles

- [hdl:cnri.dlib/magazine](http://hdl.handle.net/10338.dmlcz/141708) , <http://hdl.handle.net/10338.dmlcz/141708>
- Implementation of handle concept (Kahn-Wilensky)  
(**CNRI** – Corporation for National Research Initiatives, USA, since 1994)
- Used by DSpace (repository software), DOI (identifier), and many other...
- Main features:
  - Independent of the URN concept
  - Resolvable (own resolution system independent of DNS used by URL/URN)
    - Either a **direct resolution** by plug-in in the www-browser  
<a href="hdl:10.1045/january99-bearman">
    - or **indirect resolution** using URL-proxy  
<a href="http://hdl.handle.net/10.1045/january99-bearman">
  - <http://www.handle.net/>



## 3.7 DOI – Digital Object Identifier

The most successful identification system for the digital environment today

- **DOI:10.1006/123456**
- Initiated by the Association of American Publishers
- Built on handles technology
- Self-financing system – open, but not free (DOI allocation fee)
- System for identifying any entities (books, articles, research data, ...)
- In operation since 2000, ISO standard since 2012
- 140 million allocated DOIs, over 20,000 institutions involved
- **RA:** International DOI Foundation
- **RAGs:** 10 currently (**CrossRef** – scientific articles, DataCite – research datasets)

<http://www.doi.org> , <http://www.crossref.org>



## 3.7 DOI – Digital Object Identifier

- [DOI:10.1006/123456](https://doi.org/10.1006/123456)
- [doi:10.1000/ISBN-1-900512-44-0](https://doi.org/10.1000/ISBN-1-900512-44-0)
- [doi:10.5817/AM2013-1-17](https://doi.org/10.5817/AM2013-1-17)
- structure:
  - prefix (globally unique, assigned to registering organization by a RAG)
  - suffix (locally unique string assigned by the RO)
- DOI metadata – to be filled-in when registering DOI number in RAG Register

### References:

- [1] Balinsky, A., Ryan, J.: *Some sharp  $L^2$  inequalities for Dirac type operators*. SIGMA, Symmetry Integrability Geom. Methods Appl. (2007), 10, paper 114. [DOI 10.1037/9871-3520.5.98](https://doi.org/10.1037/9871-3520.5.98)
- [2] Brackx, F., Delanghe, R.: *Clifford Analysis*. Pitman, London, 1982. [DOI 10.2100/BDC398762](https://doi.org/10.2100/BDC398762)
- [3] Bureš, J., Sommen, F., Souček, V., Van Lancker, P.: *Rarita-Schwinger type operators in Clifford analysis*. J. Funct. Anal. 185 (2) (2001), 425–455. [DOI 10.1105/j-funct-anal.2001.425](https://doi.org/10.1105/j-funct-anal.2001.425)

# 4. Metadata



# 4. Metadata

## 1.1 Introduction

## 4.2 Classic Library Metadata

MARC, UNIMARC

## 4.3 Digital Metadata

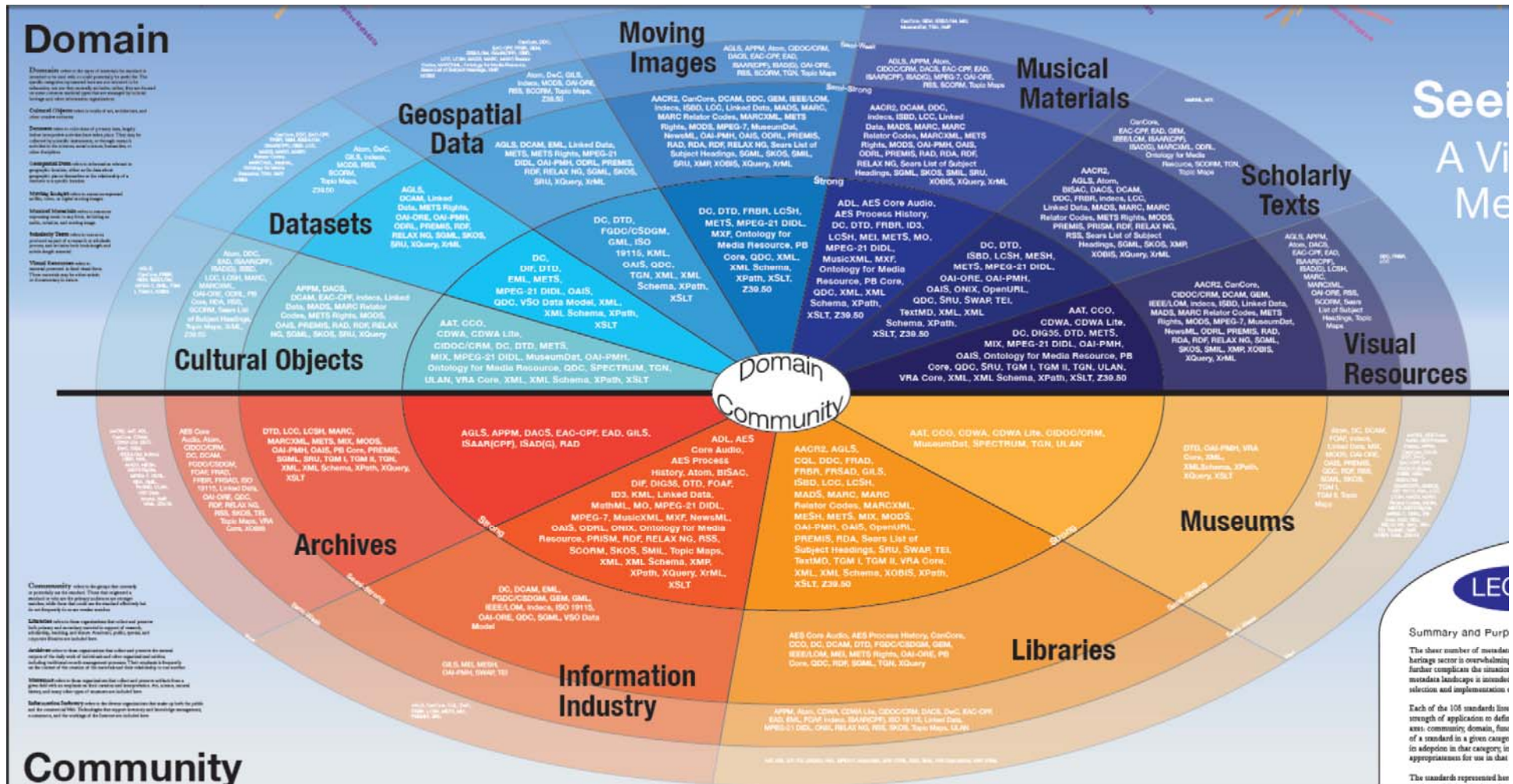
Dublin Core, MODS, METS, RDF, ...



# 4.1 Metadata – introduction

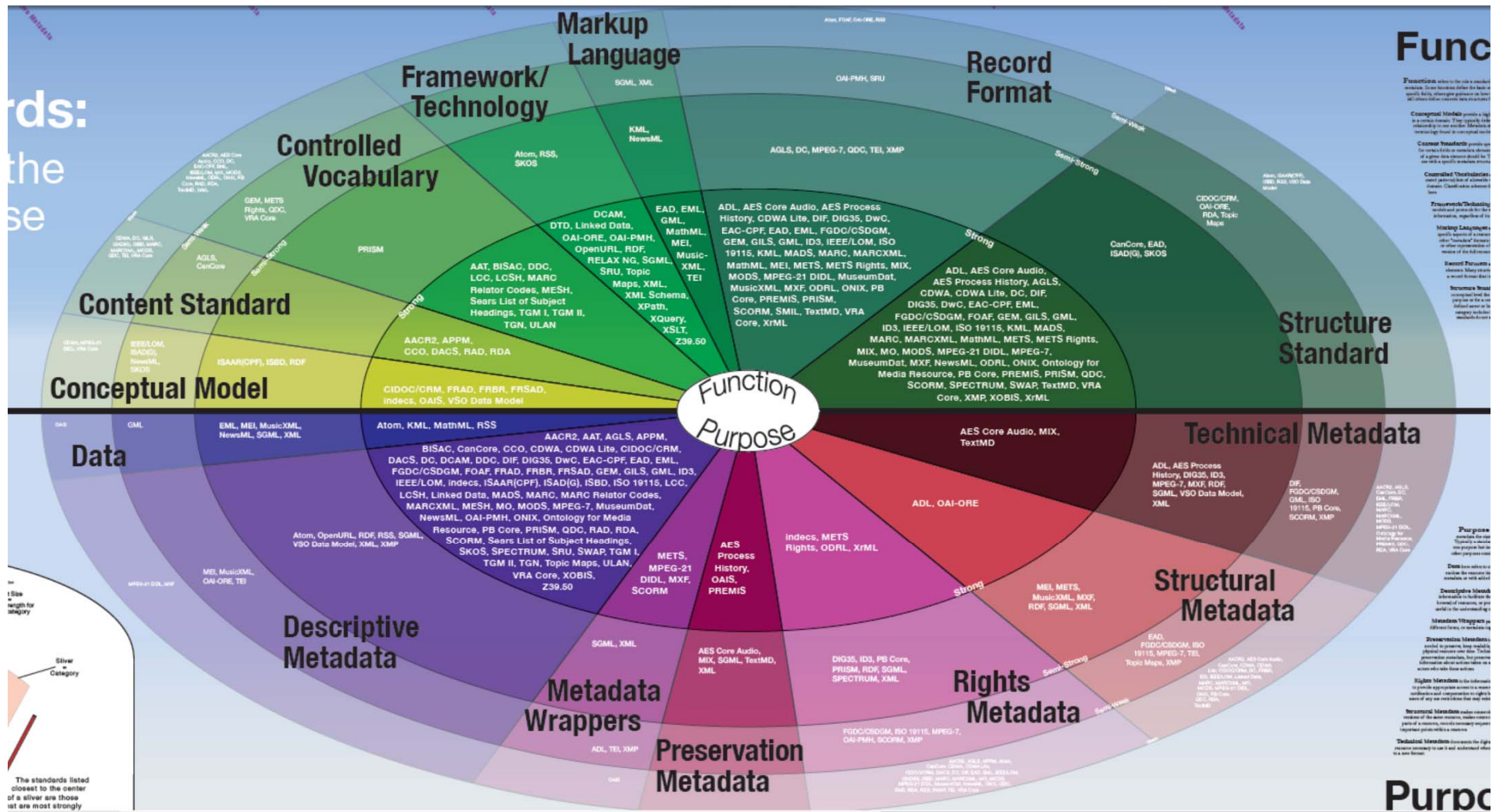
- **Metadata = (structured) data about resources**
- Metadata consists of statements we make about resources to help us find, identify, use, manage, evaluate, and preserve them
- 3 basic categories of metadata
  - **Descriptive** – resource description (to find, identify, evaluate): MARC, Dublin Core, MODS,
  - **Administrative** – resource management (technical, administrative, preservation, rights management, ...): PREMIS
  - **Structural** – resource internal structure (parts, hierarchy): METS, RDF
- **Metadata schema** (standard) – selected set of metadata elements with a defined meaning for use in a particular area (MARC, Dublin Core, TEI, MODS, MADS, RDF, Premis, ...)
- **XML** – Markup language (encoding structured documents, e.g. metadata records)

# Metadata Typology – Domains & Communities



<http://jennriley.com/metadatamap/>

# Metadata Typology – Functions & Purposes



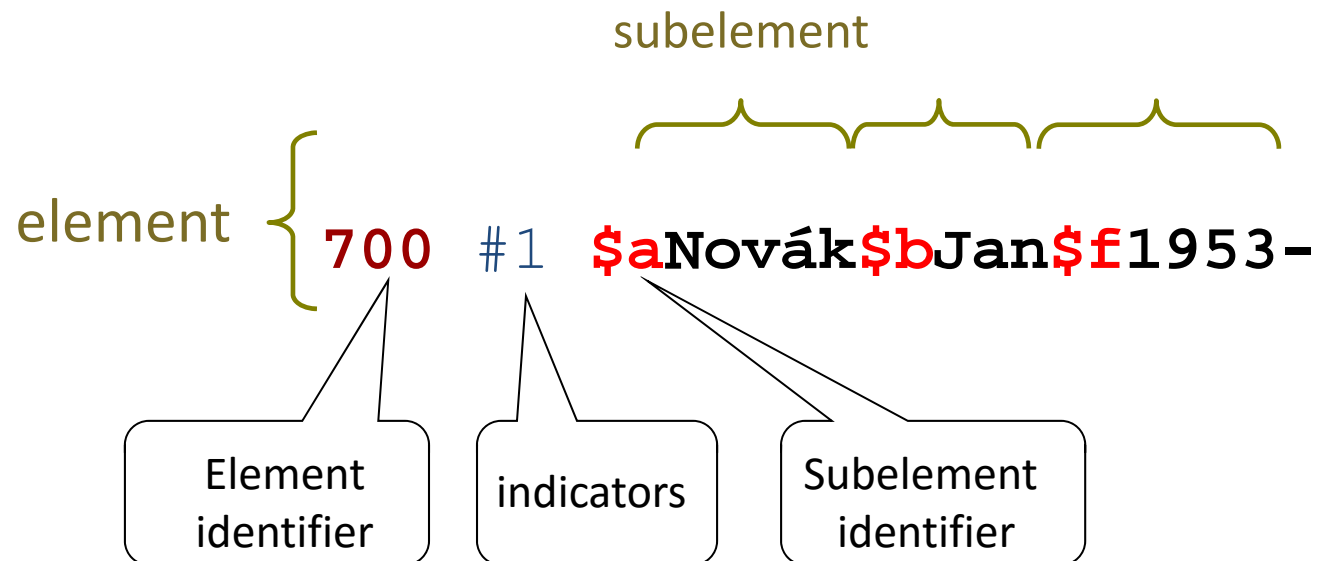
<http://jennriley.com/metadatamap/>

## 4.2 MARC Standard Family

- **MARC = MACHine Readable Cataloguing record** (Library of Congress, 1965)
- General structure of the bibliographic record (descriptive metadata for library materials – books, serials, audio, video, authorities)
  - Internal format in Library management systems
  - Exchange format for transfer of records between LMSs
- Widely used – collaboration between libraries and different systems
  - Record exchange
  - Union catalogues
- **Very reach structure** (hundreds of elements and subelements!)
- The whole family of MARC-based standards:
  - **USMARC**, **CANMARC**, **UKMARC**, ... -> **MARC21**
  - **UNIMARC** (IFLA, 1977, first as a bridge between MARCs, later as a full independent format)

## 4.2 MARC element/subelement

- MARC record consists of variable length elements
- Each element may be subdivided in subfields (with repetitioning)



## 4.2 Example – UNIMARC record

```
001    CASLIN0000001
005    19960312
010    $a80-7050-237-1
100    $a19960305d1996###k##y0czy0103###ba
101 0# $acze
102    $aCZ
200 1# $aZáznam pro souborný katalog$eUNIMARC$iTištěné monografie
      $fPracovní skupina CASLIN pro standardizaci a jmenné ...
205    $a1. vyd.
210    $aPraha$cNárodní knihovna České republiky$d1996
215    $a31 s.
225 1# $aStandardizace$vč. 4
675    $a025.3$9v
711 02 $aCASLIN$bPracovní skupina pro standardizaci a ...
801 #0 $aCZ$bABA001$c19960312$gAACR2$91
801 #3 $aCZ$bABA001$c19960515
910    $aABA001
```

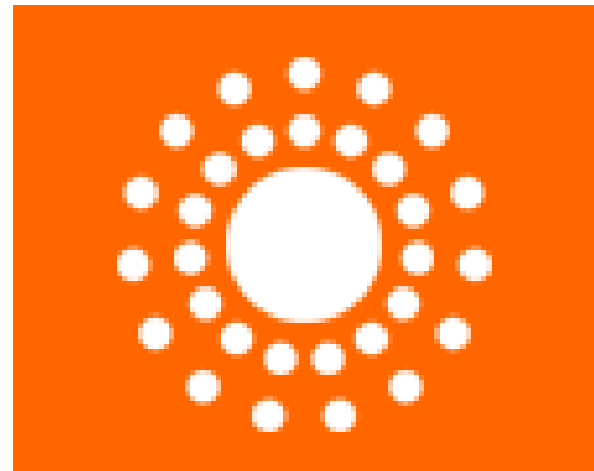
## 4.2 MARC Summary

- Detailed cataloguing rules – AACR2, RDA (how to use the format)
- Sophisticated set of tools (LCSH, authority files, ...)
- Fragmentation into many format variations
- UNIMARC – more advanced, MARC – more successful (LoC)
- Systematic development (responses to changes)
- Hundreds of millions of existing MARC records worldwide (OCLC WorldCat – 34 mil) – great legacy/burden
- Expensive creation of records (50-100 USD/record), only for highly qualified users
- Very successful but complex format -- too complicated for wider use!
- **For most applications, we need something more simple**

## 4.3 Dublin Core

Motto:

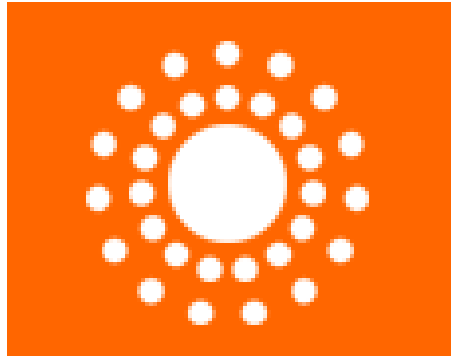
- "The association of standardized descriptive metadata with networked objects has the potential for substantially improving resource discovery capabilities by enabling field-based (e.g., author, title) searches, permitting indexing of non-textual objects, and allowing access to the surrogate content that is distinct from access to the content of the resource itself."  
(Weibel and Lagoze, 1997)
- MARC maximalist approach
- DC minimalist approach
  - **Simple**  
(core for description of resources on the Web)
  - **Universal**  
(for any kind of resources)
  - **Easy to use**  
(selfcataloguing by web users)







## 4.3 Dublin Core – 15 elements



<i>content</i>	<i>ownership</i>	<i>instantiation</i>
<b>Title</b>	<b>Creator</b>	<b>Identifier</b>
<b>Subject</b>	<b>Contributor</b>	<b>Date</b>
<b>Description</b>	<b>Publisher</b>	<b>Language</b>
<b>Coverage</b>	<b>Rights</b>	<b>Format</b>
<b>Type</b>		
<b>Source</b>		
<b>Relation</b>		
<b>Audience</b>	<b>(Provenance, Rights Holder)</b>	

- Each element is optional, repeatable, on the order of the elements does not matter
- General semantics given for each element (Title = name given to a resource)
- Syntax not given by the standard (recommendations – XML, HTML-heading,...)
- Qualified version of DC for more precise description
  - Creator.Illustrator, Date.Created, Date.Updated, Subject.Abstract
  - Date=1994-04-12:ISO8601, Subject=5.34:UDC



## 4.3 Dublin Core in HTML

```
<html>
<head>
<title> Guidance on expressing the Dublin Core within the RDF </title>
<link rel = "schema.DC" href = "http://purl.org/DC/elements/1.0/">
<meta name = "DC.Title" lang = "en" content = "Guidance on expressing...">
<meta name = "DC.Title" lang = "de" content = "Dublin Core in RDF: Eine...">
<meta name = "DC.Creator" content = "Miller, Eric">
<meta name = "DC.Creator" content = "Miller, Paul">
<meta name = "DC.Creator.Illustrator" content = "Brickley, Dan">
<meta name = "DC.Description.Abstract" content = "This paper describes work...">
<meta name = "DC.Date.Created" scheme = "ISO8601" content = "1999-07-01">
<meta name = "DC.Format" content = "text/html">
<meta name = "DC.Language" content = "en">
<meta name = "DC.Type" content = "working draft">
<meta name = "DC.Subject" scheme = "LCSH" content = "Dublin Core; DC; RDF; XML">
</head>
<body> ... </body>
</html>
```

## 4.4 MODS

- **Metadata Object Description Schema** (LoC, 2002)
- Compromise between MARC complexity and DC simplicity (19 top-elements, 64 optional subelements)
- More accurate and more modern syntax (defined as a XML schema)
- Granularity and Extensibility (the level of detail in description; embedding of sub-resource description into XML tree)
- Set of tools (MADS – Metadata Authority Description Schema)

## 4.4 MODS – 19 top elements

[titleInfo](#)

[name](#)

[typeOfResource](#)

[genre](#)

[originInfo](#)

[language](#)

[physicalDescription](#)

[abstract](#)

[tableOfContents](#)

[targetAudience](#)

[note](#)

[subject](#)

[classification](#)

[relatedItem](#)

[identifier](#)

[location](#)

[accessCondition](#)

[extension](#)

[recordInfo](#)

**TitleInfo**

- title
- subTitle
- PartNumber
- partName
- nonSort

Element attributes:

**lang, script, transliteration, ...**

## 4.4 Example – MODS record

```

<mods version="3.0">
  <titleInfo>
    <title>Hiring and recruitment in academic libraries :</title>
    <subTitle>The User Guide</subTitle> </titleInfo>
  <name type="personal">
    <namePart type="family">Raschke</namePart>
    <namePart type="given">Gregory K.</namePart>
    <displayForm>Gregory K. Raschke</displayForm> </name>
  <typeOfResource>text</typeOfResource>
  <genre>journal article</genre>
  <originInfo>
    <place><placeTerm type="text">Baltimore, Md.</placeTerm> </place>
    <publisher>Johns Hopkins University Press</publisher>
    <dateIssued>2003</dateIssued> </originInfo>
  <language>
    <languageTerm authority="iso639-2b">eng</languageTerm> </language>
  <physicalDescription>
    <form authority="marcform">print</form> <extent>15 p.</extent>
  </physicalDescription>
  ...
</mods>

```

## 4.4 MARC – MARCXML – MODS

### MARC

```
[245] 10 $aHelsinki :$ba cultural and literary history /$cNeil Kent
```

### MARCXML

```
<datafield tag="245" ind1="1" ind2="0">  
  <subfield code="a">Helsinki</subfield>  
  <subfield code="b">a cultural and literary history</subfield>  
  <subfield code="c">Neil Kent</subfield>  
</datafield>
```

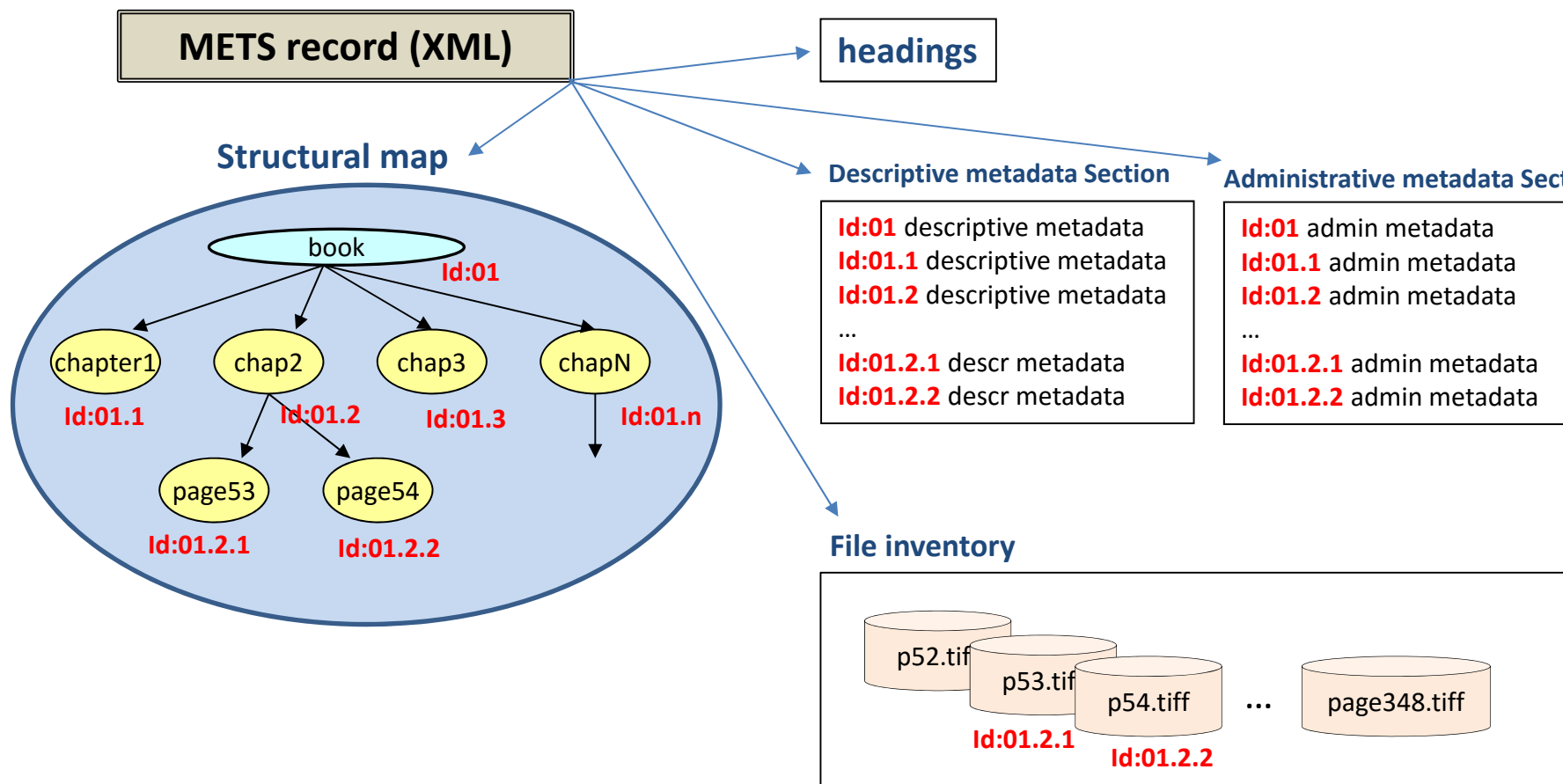
### MODS

```
<titleInfo>  
  <title>Helsinki</title>  
  <subTitle>a cultural and literary history</subTitle>  
</titleInfo>  
<note type="statement of responsibility">Neil Kent</note>
```

## 4.5 METS

- **Metadata Encoding and Transmission Standard** (LoC, 2001)
- Standard for exchanging digital objects between repositories (DLs)
- An XML schema that packs into one "package" all components of a complex DO:
  - the internal structure of the object
  - metadata (descriptive, admin, technical, etc.) for the object and all its components
  - source files that comprise the object
- The package can be moved and easily integrated into the new repository
- Example: object = one academic journal
  - Complex internal structure: Journal – Volumes – Issues - Articles
  - Thousands of metadata records for all components
  - Thousands of source files (articles)

# 4.5 METS – diagram





# 4.6 Other metadata schemas

## TEI – Text Encoding Initiative

- XML standard for marking documents and linguistic texts of any kind (books, articles, poems, dramas, ...) (1987)
- Very extensive (2000 pages)
- TEI-lite

```
<anthology>
  <poem>
    <heading>The SICK ROSE</heading>
    <stanza>
      <line>O Rose thou art sick.</line>
      <line>The invisible worm,</line>
      <line>That flies in the night</line>
      <line>In the howling storm:</line>
    </stanza>
    <stanza>
      <line>Has found out thy bed</line>
      <line>Of crimson joy:</line>
      <line>And his dark secret love</line>
      <line>Does thy life destroy.</line>
    </stanza>
  </poem>
  <!-- more poems go here -->
</anthology>
```

## RDF – Resource Description Framework

- W3C standard for describing resources on the web using simple machine-readable (understandable) statements – triplets
- Subject – predicate – object

Hamlet – Author – Shakespeare  
Hamlet – Type – tragedy  
Hamlet – Date – 1959  
Shakespeare – Nationality – British  
Shakespeare – Occupation - Writer

To be continued...

**END OF PART 1**