

From traditional libraries to digital libraries

Vittore Casarosa
University of Parma and
ISTI-CNR, Pisa, Italy

NMPLIS Summer School
5-16 July 2010, Tbilisi, Georgia

The art of librarianship

“What can be more easy (those lacking understanding say), having looked at the title pages than to write down the titles ? But these inexperienced people, who think making an index of their own few private books a pleasant task of a week or two, have no conception of the difficulties that arise or realize how carefully each book must be examined when the library numbers myriad of volumes.

In the colossal labor, which exhausts both body and soul, of making into an alphabetical catalog a multitude of books gathered from every corner of the earth there are many intricate and difficult problems that torture the mind.”

(Thomas Hyde, Bodleian Library, Oxford, 1674)

Classification and Cataloguing

- Classification means to bring related items together. Conventional libraries, in order to stack books on related subjects together, have used library classification. This facilitates the browsing approach of the information seekers
- Cataloguing creates “document surrogates”, i.e. a description of a document (a catalog record), to be used (to a certain extent) in the place of the document. Catalog records provide searching facility by Authors, Titles, Subjects, Series and other elements. Classification use notation symbols, and Authority Files provide recommended vocabularies

Library of Congress (up to 1812)

- Sacred history
- Ecclesiastical history
- Civil history
- Geography, travels
- Law
- Ethics
- Logic, rethoric, criticism
- Dictionaries, grammars
- Politics
- Trade, commerce
- Military and naval tactics
- Agriculture
- Natural history
- Medicine, surgery, chemistry
- Poetry, drama, fiction
- Arts, sciences, miscellaneous
- Gazettes (newspapers)
- Maps

In 1812 the Library of Congress moved to a “new” classification scheme, with 44 categories, and then, in 1897, to the present scheme

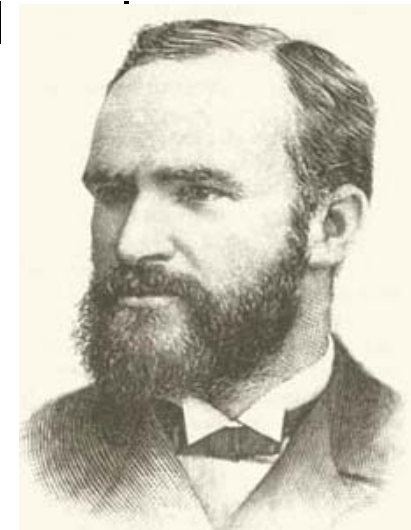
Library of Congress (today)

- A - General Works
- B - Philosophy, Psychology, Religion
- C - Auxiliary Sciences of History
- D - History: General & Outside the Americas
- E - History: United States
- F - History: United States Local & America
- G - Geography, Anthropology, Recreation
- H - Social Sciences
- J - Political Science
- K - Law
- L - Education
- M - Music
- N - Fine Arts
- P - Language and Literature
- Q - Science
- R - Medicine
- S - Agriculture
- T - Technology
- U - Military Science
- V - Naval Science
- Z - Library Science & Information Resources

DDC - Dewey Decimal System

- 000 – Computer science, information, and general
- 100 – Philosophy and psychology
- 200 – Religion
- 300 – Social sciences
- 400 – Languages
- 500 – Science and Mathematics
- 600 – Technology and applied science
- 700 – Arts and recreation
- 800 – Literature
- 900 – History and geography and biography

**DDC is presently used in over 200.000
libraries worldwide**



**Melvil Dewey
(1851-1931)
invented DDC
in 1876**

Google categories

- **Arts**
 - Movies, Music, Television, ...
- **Business**
 - Industries, Finance, Jobs, ...
- **Computers**
 - Hardware, Internet, Software, ...
- **Games**
 - Board, Roleplaying, Video, ...
- **Health**
 - Alternative, Fitness, Medicine, ...
- **Home**
 - Consumers, Homeowners, Family, etc...
- **Kids and Teens**
 - Computers, Entertainment, School, etc...
- **News**
 - Media, Newspapers, Current Events, ...
- **Recreation**
 - Food, Outdoors, Travel, ...
- **Reference**
 - Education, Libraries, Maps, ...
- **Regional**
 - Asia, Europe, North America, etc...
- **Science**
 - Biology, Psychology, Physics, etc...
- **Shopping**
 - Autos, Clothing, Gifts, ...
- **Society**
 - Issues, People, Religion, ...
- **Sports**
 - Basketball, Football, Soccer, ...

Yahoo categories

- **Automotive**
 - Dealers, Motorcycles, Repair
- **Community**
 - Government, Libraries, Religion
- **Computers and Internet**
 - Internet Services, Computer Stores etc...
- **Education and Instruction**
 - Colleges and Universities, K-12, Adult Education etc...
- **Entertainment and Arts**
 - Movies, Music, Nightclubs etc...
- **Food and Dining**
 - Catering, Grocers, Restaurants etc...
- **Health and Medicine**
 - Doctors, Hospitals, Fitness etc...
- **Home and Garden**
 - Housewares, Nurseries, Utilities etc...
- **Legal and Financial**
 - Banks, Insurance, Law Firms etc...
- **Other Professional Services**
 - Printing, Pets, Jobs, Cellular etc...
- **Personal Care**
 - Drug Stores, Hair Care, Nail Salons etc...
- **Real Estate**
 - Rental Agencies, Apartment Complexes, Agents etc...
- **Recreation and Sports**
 - Golf, Hobbies, Sporting Goods etc...
- **Retail Shopping**
 - Books, Malls, Flowers etc...
- **Travel and Transportation**
 - Hotels, Travel Agents, Taxis etc...
- **Business to Business**
 - Manufacturing, Business Services, Construction etc...

YAHOO: Yet Another Hierarchical Organization

Bibliographic Elements

- After classifying an item, we have to define the “bibliographic elements” that will allow to find it, and traditionally the main elements are:
 - Titles
 - Authors
 - Subjects (subject headings)
- Need for “Authority files”
 - Name Authority files
 - (e.g. 47 different spellings for the name of Muammar Qaddafi)
 - Library of Congress Subject Headings (LCSH)
 - About 2 million of cross-referenced entries (60% of which are preferred terms)

Need for Authority Files

Table 2.1 Spelling variants of the name Muammar Qaddafi.

Qaddafi, Muammar	Muammar al-Qadhafi	Qathafi, Muammar
Gadhafi, Mo ammar	Mu ammar al-Qadhdhafi	Gheddafi, Muammar
Kaddafi, Muammar	Qadafi, Mu ammar	Muammar Gaddafy
Qadhafi, Muammar	El Kazzafi, Moamer	Muammar Ghadafi
El Kadhafi, Moammar	Gaddafi, Moamar	Muammar Ghaddafi
Kadhafi, Moammar	Al Qathafi, Mu ammar	Muammar Al-Kaddafi
Moammar Kadhafi	Al Qathafi, Muammar	Muammar Qathafi
Gadafi, Muammar	Qadhdhafi, Mu ammar	Muammar Gheddafi
Mu ammar al-Qadafi	Kaddafi, Muammar	Khadafy, Moammar
Moamer El Kazzafi	Muammar al-Khaddafi	Qudhafi, Moammar
Moamar al-Gaddafi	Mu amar al-Kad'afi	Qathafi, Mu'Ammar el
Mu ammar Al Qathafi	Kad'afi, Mu amar al-	El Qathafi, Mu'Ammar
Muammar Al Qathafi	Gaddafy, Muammar	Kadaffi, Momar
Mo ammar el-Gadhafi	Gadafi, Muammar	Ed Gaddafi, Moamar
Muammar Kaddafi	Gaddafi, Muammar	Moamar el Gaddafi
Moamar El Kadhafi	Kaddafi, Muamar	

Bibliographic records

- “The art of cataloguing” has been developed by librarians over centuries
- Up to the fifties and sixties the bibliographic records were materialized in **catalog cards**
 - a catalog was comprised of many set of identical cards, each set ordered by a different **access point**
- Starting with the early seventies, the diffusion of computers brought the catalogs on line
- The problem of **interoperability** started to become apparent

OPAC

On-line Public Access Catalog (early seventies)

- **Images** of the catalog cards and/or **text** contained in the catalog cards
- Many advantages over traditional access via physical catalog cards
 - More than one access point
 - Author=Salton, Gerald AND Title=Modern Library Services
 - Author=Salton AND Title=Library
 - Any=Library
- WorldCat (maintained by OCLC – Online Computer Library Center) is presently the biggest OPAC
 - started in 1971
 - about 10 thousand libraries from more than 90 countries
 - more than 90 million records
 - 1200 million physical and digital assets
 - 360 languages

MARC21

MAchine Readable Cataloging

- Started in the late sixties and developed by the Library of Congress to facilitate catalog sharing
- Provides a machine readable representation of a catalog card
- Based on a system of numbers, letters and symbols to identify fields in the record
- Provides a precise (sharable) description of the object
- Many “national” versions (UKMARC, CANMARC, AUSMARC, DANMARC, ANNAMARC, INTERMARC, etc)
- UNIMARC (Universal MARC) as standard format for exchange of information
 - e.g. USMARC to UNIMARC to AUSMARC

Catalog card

- Name: Arnosky, Jim.
- Title proper: Raccoons and ripe corn.
- Statement of responsibility: Jim Arnosky
- Edition statement: 1st ed.
- Place of publication: New York
- Name of publisher: Lothrop, Lee & Shepard Books
- Date of publication: c1987
- Pagination: 25 p.
- Illustrative matter: col. ill.
- Size: 26 cm
- Summary: Hungry raccoons feast at night in a field of ripe corn
- Topical subject: Raccoons
- Local call number: 599.74 ARN
- Local barcode number: 8009
- Local price: \$15.00

MARC representation

"SIGNPOSTS"			DATA
100	1#	\$a	Arnosky, Jim.
245	10	\$a \$c	Raccoons and ripe corn / Jim Arnosky.
250	##	\$a	1st ed.
260	##	\$a \$b \$c	New York : Lothrop, Lee & Shepard Books, c1987.
300	##	\$a \$b \$c	25 p. : col. ill. ; 26 cm.
520	##	\$a	Hungry raccoons feast at night in a field of ripe corn.
650	#1	\$a	Raccoons.
900	##	\$a	599.74 ARN
901	##	\$a	8009
903	##	\$a	\$15.00

MARC structure

- Each bibliographic record is divided logically into fields. There is a field for the author, a field for title information, and so on. These fields are subdivided into one or more "subfields." Many fields are repeatable
- Each field is associated with a 3-digit number called a "*tag*". A tag identifies the field (i.e. the kind of data) that follows.
- Two character positions follow each tag (with the exception of Fields 001 through 009). One or both of these character positions may be used for "*indicators*". In some fields, only the first or second position is used; in some fields, both are used; and in some fields neither is used. Each indicator value is a number from 0 to 9. Empty indicators are marked by a "placeholder".
- Most fields contain several related pieces of data. Each type of data within the field is called a "*subfield*", and each subfield is preceded by a subfield code. Subfield codes are one lowercase letter (occasionally a number) preceded by a delimiter.

Actual MARC record

```
01041cam 2200265 a 450000100200000000300040002000
50017000240080041000410100024000820200025001060200
04400131040001800175050002400193082001800217100003
20023524500870026724600360035425000120039026000370
04023000029004395000042004685200220005106500033007
30650001200763^###89048230#/AC/r91^DLC^19911106082
810.9^891101s1990####maua###j#####000#0#eng##^##$
a###89048230#/AC/r91^##$a0316107514 :$c$12.95^##$a
0316107506 (pbk.) :$c$5.95 ($6.95 Can.)^##$aDLC$cD
LC$dDLC^00$aGV943.25$b.B74 1990^00$a796.334/2$220^
10$aBrenner, Richard J.,$d1941-^10$aMake the team.
$pSoccer :$ba heads up guide to super soccer! /$cR
ichard J. Brenner.^30$aHeads up guide to super soc
cer.^##$alst ed.^##$aBoston :$bLittle, Brown,$ccl19
90.^##$a127 p. :$bill. ;$c19 cm.^##$a"A Sports ill
ustrated for kids book."^##$aInstructions for impr
oving soccer skills. Discusses dribbling, heading,
playmaking, defense, conditioning, mental attitud
e, how to handle problems with coaches, parents, a
nd other players, and the history of soccer.^#0$aS
occer$vJuvenile literature.^#1$aSoccer.^\\
```

Leader record
(24 digits)

ID of the tag field
(3 digits)

Length of the field
(4 digits)

Position of the starting
character of the field
(5 digits)

Display of a catalog card

TITLE: Make the team. Soccer : a heads up guide to super soccer! / Richard J. Brenner.

ADDED TITLE: Heads up guide to super soccer

AUTHOR: Brenner, Richard J., 1941-

PUBLISHED: 1st ed. Boston : Little, Brown, c1990.

MATERIAL: 127 p. : ill. ; 19 cm.

NOTE: "A Sports illustrated for kids book."

NOTE: Instructions for improving soccer skills.
Discusses dribbling, heading, playmaking, defense, conditioning, mental attitude, how to handle problems with coaches, parents, and other players, and the history of soccer.

SUBJECT: Soccer--Juvenile literature.
Soccer.

Copies Available: GV943.25 .B74 199

Metadata: some definitions

- Machine-understandable information about [Web resources or other things](#) (Tim Berners-Lee 1997)
- Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics; a user might be a program or a person (Lorcan Dempsey 1998)
- Structured data about [resources](#) that can be used to help support a wide range of operations (Michael Day, 2001)
- Structured data about data (DCMI 2003)
- Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an [information resource](#) (NISO 2004)

Different use of the term “metadata”

- Database Management Systems (schemas of relational databases): machine understandable information
- World Wide Web (since the mid-1990's): records that describe “resources”
- Library environment: any formal scheme of “document” description, applying to any type of object, digital or non-digital
 - therefore, traditional library cataloging is a form of metadata; MARC 21 and the rule sets used with it, such as AACR2R (Anglo American Cataloguing Rules, Release 2), are metadata standards
 - other metadata schemes have been developed to describe various types of textual and non-textual objects including published books, electronic documents, archival finding aids, art objects, educational and training materials, scientific datasets etc.

Another point of view

“Then there is the question of cataloguing and metadata.

My view of the latter is that it is an ill-considered attempt to find some kind of Third Way between the wilderness of search engines and free text searching and the grand architecture of bibliographic control that librarians have developed over the last 150 years.

I think that metadata is the product of those with no knowledge of, or regard for, cataloguing; they are bibliographic alchemists seeking the philosopher’s stone that will offer us effective cataloguing without expense and effective access without controlled vocabularies.

There is no such thing and the sooner that notion is disposed of, the better”

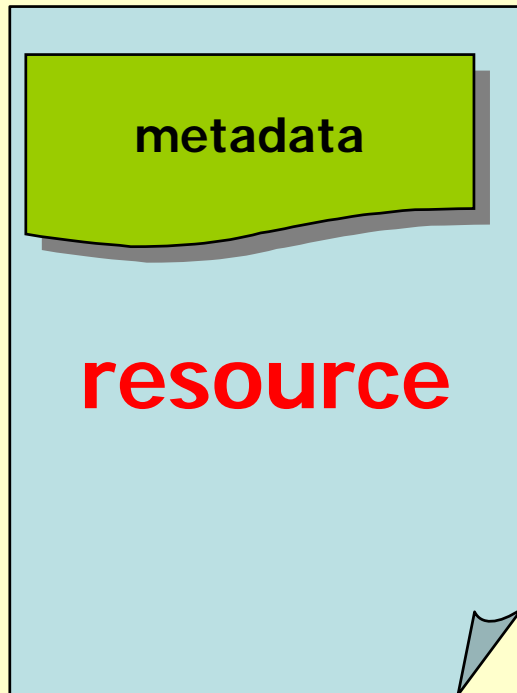
(Michael Gorman, Dean of Library Services at California State University, past President of the American Library Association, November 2000)

Not simply a cataloguing record

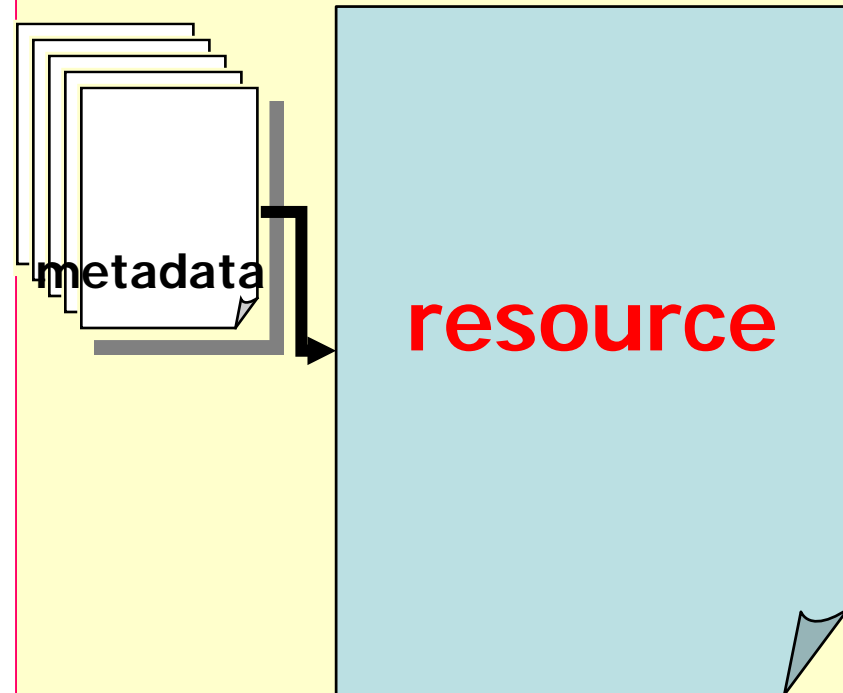
- An important reason for creating descriptive metadata is to facilitate discovery of relevant information, as it serves the same functions in resource discovery as good cataloging does by:
 - allowing resources to be found by relevant criteria
 - identifying resources
 - bringing similar resources together
 - distinguishing dissimilar resources
 - giving location information
- In addition to resource discovery, metadata can
 - help organize electronic resources
 - facilitate interoperability and legacy resource integration
 - provide digital identification
 - support archiving and preservation

Storing Metadata

embedded metadata



stand-alone metadata



Emerging new requirements (mid-90's)

- Increase in the amount of information available on-line (data bases, repositories, the Web, etc)
- Increase in the variety of information available on-line (text, sound, images, video, 3D, etc)
- Description of information not always done by "specialists"
- Scholarly publishing (open access and non-open access)
 - Self-publishing

Institutional Repositories

- Maintained by universities and research institutions for “self-publishing” (potential problems with copyright owners)
- An *e-print* is an author self-archived document. The content of an e-print is usually the result of scientific or other scholarly research.
- Repositories contain scholarly publications
 - Reports
 - Working papers
 - Pre- and post-prints of articles and books
 - Doctoral thesis
 - Data supporting research
 - References and professional databases related to research topics
- Improve visibility of research results
- Make research results freely available on the Web
- Contribute to the Open Access movement

Dublin Core Metadata Initiative

- Dublin Core
 - Dublin: Dublin, Ohio, 1995
 - Core: minimal set of broad and generic elements
- Dublin Core was originally developed with an eye to describing document-like objects
 - Descriptions easy to create (unlike MARC)
- Despite initial focus, has proved to be general enough to describe “any” type of objects
 - unlike catalog records, often tied to specific application fields
- It is now a widely used international standard
 - ISO Standard 15836-2003
 - NISO Standard Z39.85-2007
 - IETF RFC 5013

Fifteen Core Elements (1996)

Definition of **elements** (or **terms**) to describe **resources**

Content	Intellectual Property	Instantiation
Title	Creator	Date
Subject	Contributor	Format
Description	Publisher	Identifier
Type	Rights	Language
Source		
Relation		
Coverage		

Characteristics of the Dublin Core

- All elements optional
- All elements repeatable
- Elements may be displayed in any order
- International in scope
- Extensible (Qualified Dublin Core)
- Dublin Core Principles
 - Dumb-Down
 - One-to-One
 - Appropriate Values

Simple and Qualified Dublin Core

- Simple Dublin Core is limited to the original 15 elements
- Qualified Dublin Core includes, in addition:
 - New Elements
 - Qualifiers
 - Element Refinements
 - Encoding Schemes

New elements

- Over time, the initial set of 15 elements has been enriched with a small number of (carefully selected) new elements
 - Audience
 - Provenance
 - RightsHolder
 - Instructional Method
 - Accrual Method
 - Accrual Periodicity
 - Accrual Policy

Qualifiers

- Element refinements
 - These qualifiers make the meaning of an element narrower or more specific. A refined element shares the meaning of the unqualified element, but with a more restricted scope.
- Encoding schemes
 - These qualifiers identify schemes that aid in the interpretation of an element value.
 - Vocabulary encoding schemes
 - A value expressed using a vocabulary encoding scheme will be a token selected from a controlled vocabulary (e.g., a term from a classification system or set of subject headings)
 - Syntax encoding schemes
 - A value expressed using a syntax encoding scheme will be a string formatted in accordance with a formal notation (e.g., "2000-01-01" as the standard expression of a date).
 - If an encoding scheme is not understood by a client or agent, the value may still be useful to a human reader

Element refinements

DCMES Element	Element Refinement(s)
Title	Alternative
Description	Table Of Contents Abstract
Date	Created Valid Available Issued Modified Date Accepted Date Copyrighted Date Submitted
Format	Extent Medium
Identifier	Bibliographic Citation

Relation	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format Conforms To
Coverage	Spatial Temporal
Rights	AccessRights License
Audience	Mediator Education Level

Encoding schemes

DCMES Element	Element Encoding Scheme(s)
Subject	LCSH MeSH LCC DDC UDC
Date	DCMI Period W3C-DTF
Type	DCMI Type Vocabulary
Format	IMT (MIME Types)
Identifier	URI
Source	URI

Language	ISO 639-2 RFC 3066
Relation	URI
Coverage spatial	DCMI Point ISO 3166 DCMI Box TGN
Coverage temporal	DCMI Period W3C-DTF
Rights license	URI

Characteristics of the Dublin Core

- All elements optional
- All elements repeatable
- Elements may be displayed in any order
- International in scope
- Extensible
 - Element refinements
 - Encoding schemes
- **Dublin Core Principles**
 - Dumb-Down
 - One-to-One
 - Appropriate Values

The Dumb-Down principle

The fifteen core elements are usable with or without qualifiers

- Qualifiers make elements more specific
 - Element Refinements narrow meanings, never extend
 - Encoding Schemes give context to element values
- If your software encounters an unfamiliar qualifier, look it up –or **just ignore it!**

The One-to-One principle

- Describe one manifestation of a resource with one record
 - Example: a digital image of the Mona Lisa is not described as if it were the same as the original painting
- Separate descriptions of resources from descriptions of the agents responsible for those resources
 - Example: email addresses and affiliations of creators are attributes of the creator, not the resource
- If needed, gather related descriptions into a “description set” (record)

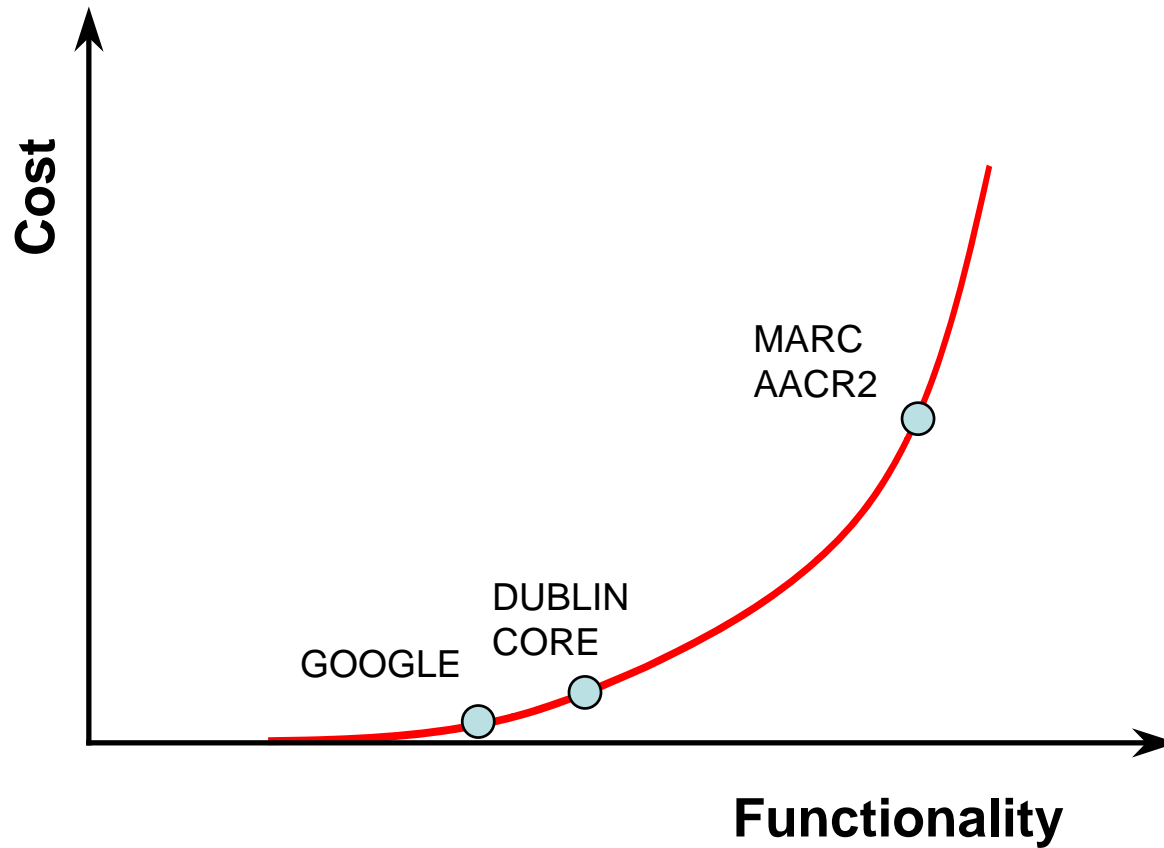
The “Appropriate Values” principle

- Use elements, element refinements and qualifiers to meet the needs of your local context, but. . .
- Remember that your metadata may be interpreted by machines and people, so. . .
- Consider whether the values you use will aid discovery outside your local context and. . .
- Make decisions about your local practices accordingly

MARC and Dublin Core

- They represent opposite ends of the complexity spectrum.
- MARC is a comprehensive, well-developed, carefully controlled scheme intended to be generated by professional catalogers for use in libraries.
- Dublin Core is an intentionally minimalist standard intended to be applied to a wide range of digital library materials by people who are not trained in library cataloging.
- These two schemes are of interest not only for their practical value, but also to highlight diametrically opposed underlying philosophies.

Defining metadata



Expressing Dublin Core

- Expressing Dublin Core in HTML/XHTML meta and link elements (<http://dublincore.org/documents/dcq-html/>)
 - Status: Recommendation (2008)
 - Note: The previous specifications (issued in 2003) still remain one of the most accessed DCMI documents/
- Guidelines for implementing Dublin Core in XML (<http://dublincore.org/documents/dc-xml-guidelines/>)
 - Status: Recommendation (2003)
 - Note: Recently (2008) a revision of this specification has been issued
- Expressing Dublin Core metadata using RDF, the Resource Description Framework (<http://dublincore.org/documents/dc-rdf/>)
 - Status: Recommendation

Example - a simple DC record

The resource being described
is the home page of UKOLN

```
<?xml version="1.0"?>
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/
    http://example.org/myapp/schema.xsd"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>
    UKOLN
  </dc:title>
  <dc:description>
    UKOLN is a national focus of expertise .....
  </dc:description>
  <dc:publisher>
    UKOLN, University of Bath
  </dc:publisher>
  <dc:identifier>
    http://www.ukoln.ac.uk/
  </dc:identifier>
</metadata>
```

Metadata standards for DLs

- Interoperability and object exchange requires the use of established standards
- Many digital objects are complex and are comprised of multiple files
- Complex digital objects require many more forms of metadata than physical objects for their management and use
 - Descriptive
 - Technical
 - Structural
 - Administrative
 - Digital provenance/events
 - Rights/Terms and conditions
- XML is the de-facto standard for metadata descriptions on the Internet

Resource Description Formats

- MARCXML
 - MARC 21 data in an XML structure
- MODS (Metadata Object Description Standard)
 - XML markup for selected metadata from existing MARC 21 records as well as original resource description
- MADS (Metadata Authority Description Standard)
 - XML markup for selected authority data from MARC21 records as well as original authority data
- EAD (Encoded Archival Description)
 - XML markup designed for encoding archival information

Digital Library Standards

- METS (Metadata Encoding & Transmission Standard)
 - Structure for encoding descriptive, administrative, and structural metadata
- MIX (NISO Metadata for Images in XML)
 - XML schema for encoding technical data elements required to manage digital image collections
- PREMIS (Preservation Metadata)
 - A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials.
- TextMD (Technical Metadata for Text)
 - XML schema that details technical metadata for text-based digital objects.
- TEI (Text Encoding Initiative)
 - Specifications of encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics.

Representation of knowledge

- Description of Information (resources) through metadata is an exercise in “knowledge representation”
- Knowledge representation might be language dependent
- Knowledge representation is the “Holy Graal” of Computer Science
 - Artificial Intelligence
 - Expert Systems
 - Ontologies
 -
- Many models/languages proposed in the last 40 years
- Most of the advances due to “brute force” methods
- Conceptual models of interest to DL: FRBR, RDF, DCAM

Functional Requirements for Bibliographic Records

- An abstract conceptual model, based on the entity-relationship model
- Entities, Relationships, Attributes
- User Tasks
 - Find
 - Identify
 - Select
 - Obtain
- Mandatory elements for a national level bibliographic record

FRBR Entities

Group 1	Work, Expression, Manifestation, Item
Group 2	Person, Corporate body
Group 3	Concept, Object, Event, Place

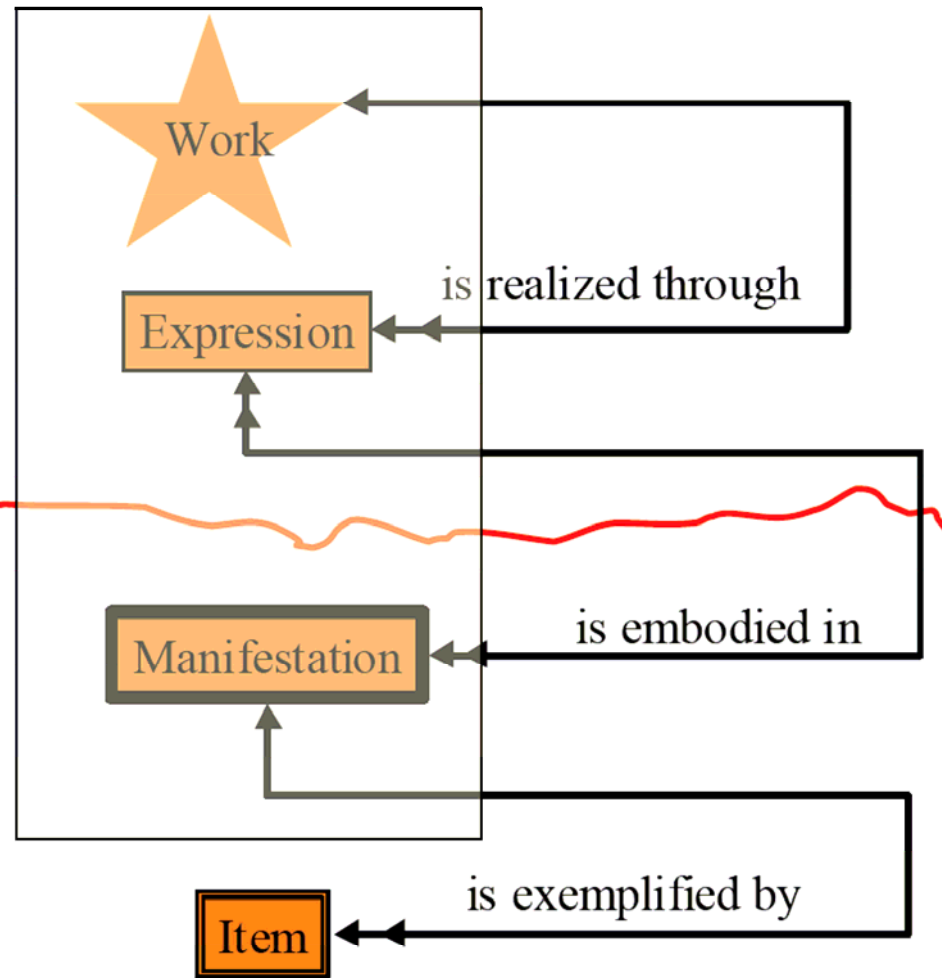
FRBR – Group 1 entities



products of
 intellectual
 or artistic
 endeavour

Intellectual/
 artistic content

Physical -
 recording of
 content



Examples of Work and Expressions

- w1 Henry Gray's Anatomy of the human body
 - e1 text and illustrations for the first edition
 - e2 text and illustrations for the second edition
 - e3 text and illustrations for the third edition
 -
- w1 J. S. Bach's The art of the fugue
 - e1 the composer's score for organ
 - e2 an arrangement for chamber orchestra by Anthony Lewis
 -
- w1 Jules et Jim (motion picture)
 - e1 the original French language version
 - e2 the original with English subtitles added
 -

Examples of <different> Works

- William Shakespeare's [Romeo and Juliet](#)
- Franco Zeffirelli's motion picture [Romeo and Juliet](#)
- Baz Lurhmann's motion picture William Shakespeare's [Romeo and Juliet](#)
-

Examples of Expressions

- Franz Schubert's Trout quintet (Work)
 - e1 the composer's notated music
 - e2 the musical work as performed by Rosina Lhevinne, piano, Stuart Sankey, double bass, and members of the Juilliard String Quartet
 - e3 the musical work as performed by Jörg Demus, piano, and the members of the Collegium Aureum
 - e4 the musical work as performed by Emanuel Ax, piano, members of the Guarneri String Quartet, and Julius Levine, double bass
 -

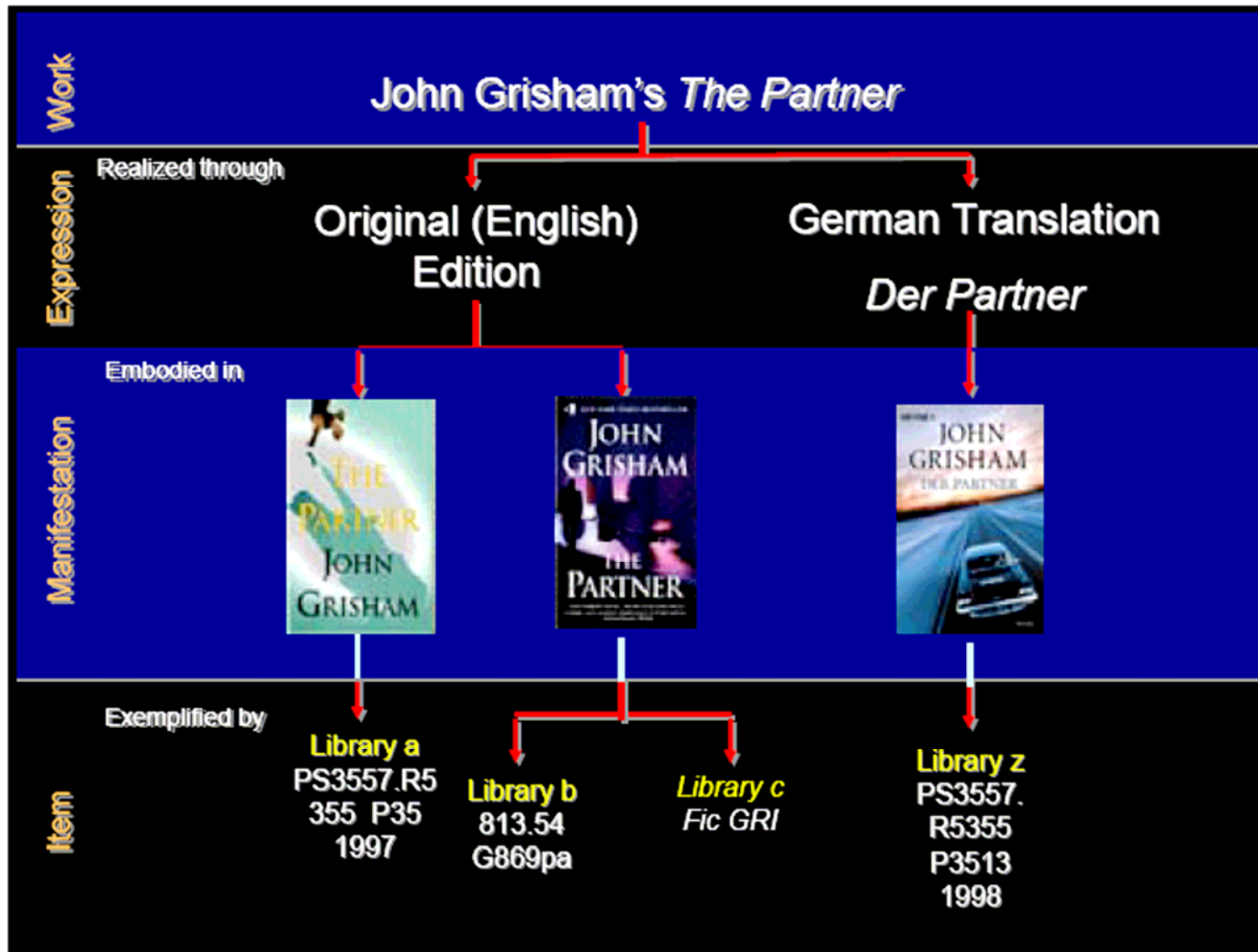
Examples of Manifestations

- J. S. Bach's Six suites for unaccompanied cello (Work)
 - e1 performances by Janos Starker recorded partly in 1963 and completed in 1965
 - m1 recordings released on 33 1/3 rpm sound discs in 1966 by Mercury
 - m2 recordings re-released on compact disc in 1991 by Mercury
 - e2 performances by Yo-Yo Ma recorded in 1983
 - m1 recordings released on 33 1/3 rpm sound discs in 1983 by CBS Records
 - m2 recordings re-released on compact disc in 1992 by CBS Records

Example of Item

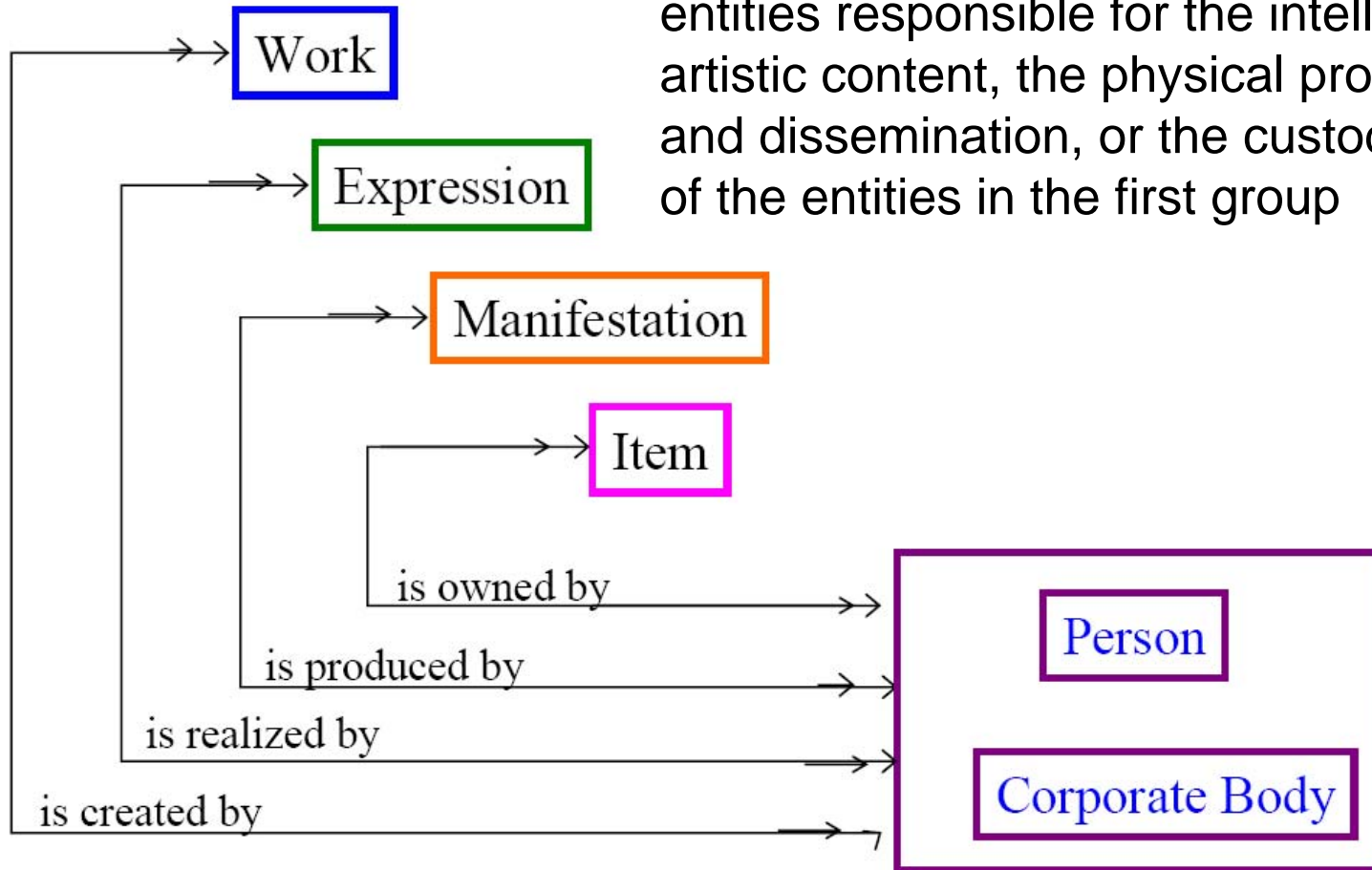
- w1 Ronald Hayman's Playback
 - e1 the author's text edited for publication
 - m1 the book published in 1973 by Davis-Poynter
 - i1 copy autographed by the author

From Work to Item

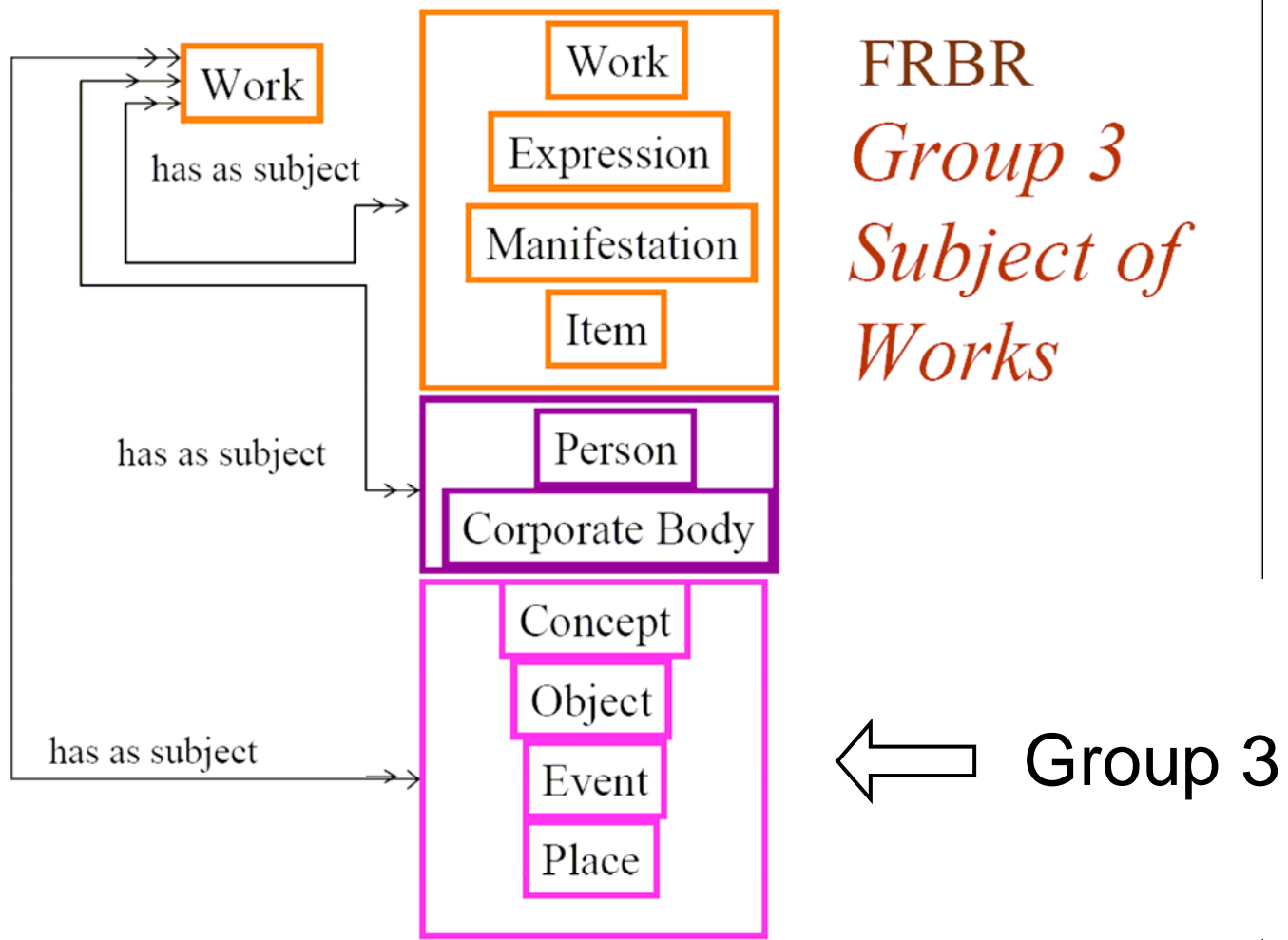


FRBR – Group 2 entities

entities responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of the entities in the first group

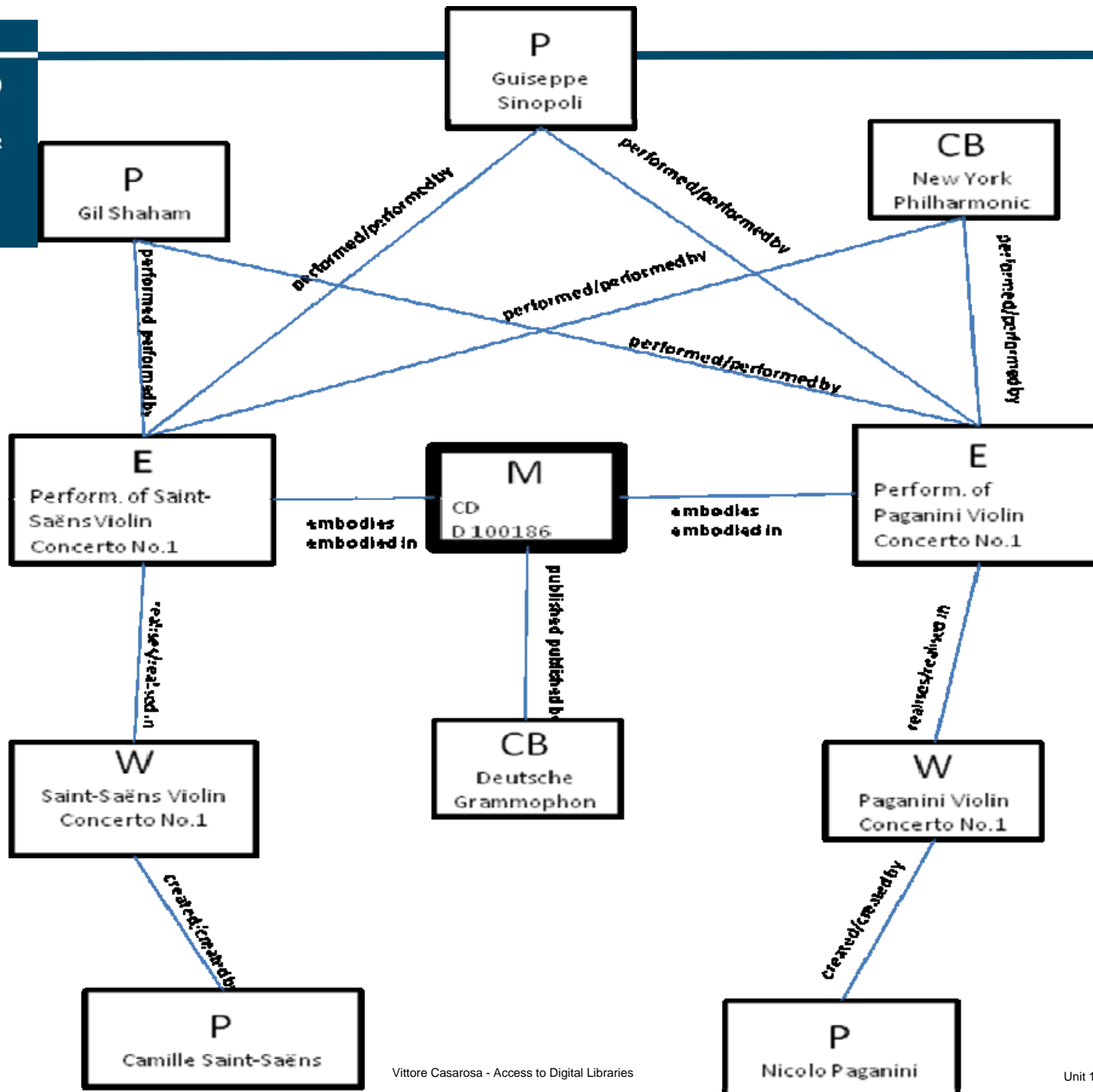


FRBR – Group 3 entities



Example of FRBR







Shakespeare, William, 1564-1616. Hamlet. French.

LC Control No. : 47023612

LCCN Permalink : <http://lccn.loc.gov/47023612>

Type of Material : Book (Print, Microform, Electronic, etc.)

Personal Name : [Shakespeare, William, 1564-1616.](#)

Main Title : ... Hamlet, traduit par André Gide.

Published/Created : [Paris] Gallimard [1946]

Description : 2 p. l., 7-237, [2] p. 17 cm.

CALL NUMBER : [PR2779.H3 G5](#)Copy 1

-- Request in : Jefferson or Adams Bldg General or
Area Studies Reading Rms



Shakespeare, William, 1564-1616. **Hamlet.** French.

LC Control No. : 47023612

LCCN Permalink : <http://lcn.loc.gov/47023612>

Work

Type of Material : Book (Print, Microform, Electronic, etc.)

Personal Name : [Shakespeare, William, 1564-1616.](#)

Main Title : ... Hamlet, traduit par André Gide.

Published/Created : [Paris] Gallimard [1946]

Description : 2 p. l., 7-237, [2] p. 17 cm.

CALL NUMBER : [PR2779.H3 G5](#)Copy 1

-- Request in : Jefferson or Adams Bldg General or
Area Studies Reading Rms



Shakespeare, William, 1564-1616. Hamlet. **French.**

LC Control No. : 47023612 **Expression**

LCCN Permalink : <http://lccn.loc.gov/47023612>

Type of Material : Book (Print, Microform, Electronic, etc.)

Personal Name : [Shakespeare, William, 1564-1616.](#)

Main Title : ... Hamlet, traduit par André Gide.

Published/Created : [Paris] Gallimard [1946]

Description : 2 p. l., 7-237, [2] p. 17 cm.

CALL NUMBER : [PR2779.H3 G5](#) Copy 1

-- **Request in** : Jefferson or Adams Bldg General or
Area Studies Reading Rms



Shakespeare, William, 1564-1616. Hamlet. French.

LC Control No. : 47023612

Manifestation

LCCN Permalink : <http://lccn.loc.gov/47023612>

Type of Material : Book (Print, Microform, Electronic, etc.)

Personal Name : [Shakespeare, William, 1564-1616.](#)

Main Title : ... Hamlet, traduit par André Gide.

Published/Created : [Paris] Gallimard [1946]

Description : 2 p. l., 7-237, [2] p. 17 cm.

CALL NUMBER : [PR2779.H3 G5](#)Copy 1

-- Request in : Jefferson or Adams Bldg General or
Area Studies Reading Rms



Shakespeare, William, 1564-1616. Hamlet. French.

LC Control No. : 47023612

LCCN Permalink : <http://lcn.loc.gov/47023612>

Type of Material : Book (Print, Microform, Electronic, etc.)

Personal Name : [Shakespeare, William, 1564-1616.](#)

Main Title : ... Hamlet, traduit par André Gide.

Published/Created : [Paris] Gallimard [1946]

Description : 2 p. l., 7-237, [2] p. 17 cm.

CALL NUMBER : [PR2779.H3 G5](#)Copy 1

-- Request in : Jefferson or Adams Bldg General or
Area Studies Reading Rms

Item
↙

The “FRBR family”

- FRBR: the original framework
 - All entities, focusing on Group 1
- FRAR (FRAD): Functional Requirements for Authority Records/Data
 - Focus on Group 2
 - Published in 2009
- FRSAR (FRSAD): Functional Requirements for Subject Authority Records/Data
 - Focus on ‘aboutness’ (Group 3)
 - In revision after IFLA review, will be published in 2010

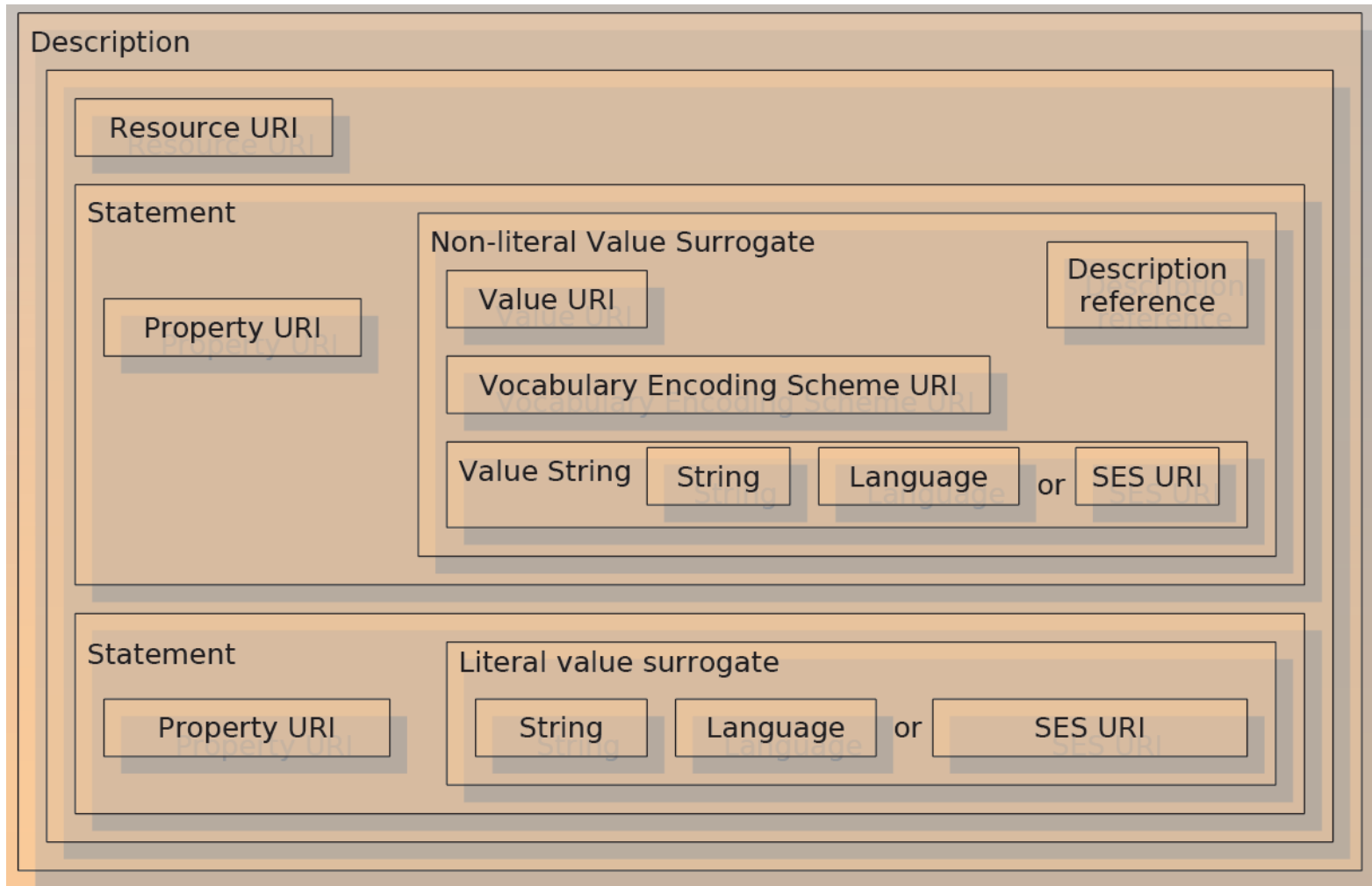
DCAM – DC Abstract Model (1/2)

- Resource
 - a resource is anything that has identity. For example, a resource may be an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources.
- Property
 - a property is a specific aspect, characteristic, attribute, or relation used to describe a resource.
- Record
 - a record is some structured metadata about a resource, comprising one or more properties and their associated values.

DCAM – DC Abstract Model (2/2)

- Dublin Core is used to make descriptions about *resources*
- A description is made up of
 - the URI of the resource being described (resource URI)
 - one or more statements (about just that one resource)
- Each statement is made up of
 - a property URI (that identifies a property)
 - a value URI (that identifies a value) and/or
 - one or more representations of the value (a “value surrogate”)

Summary of the model

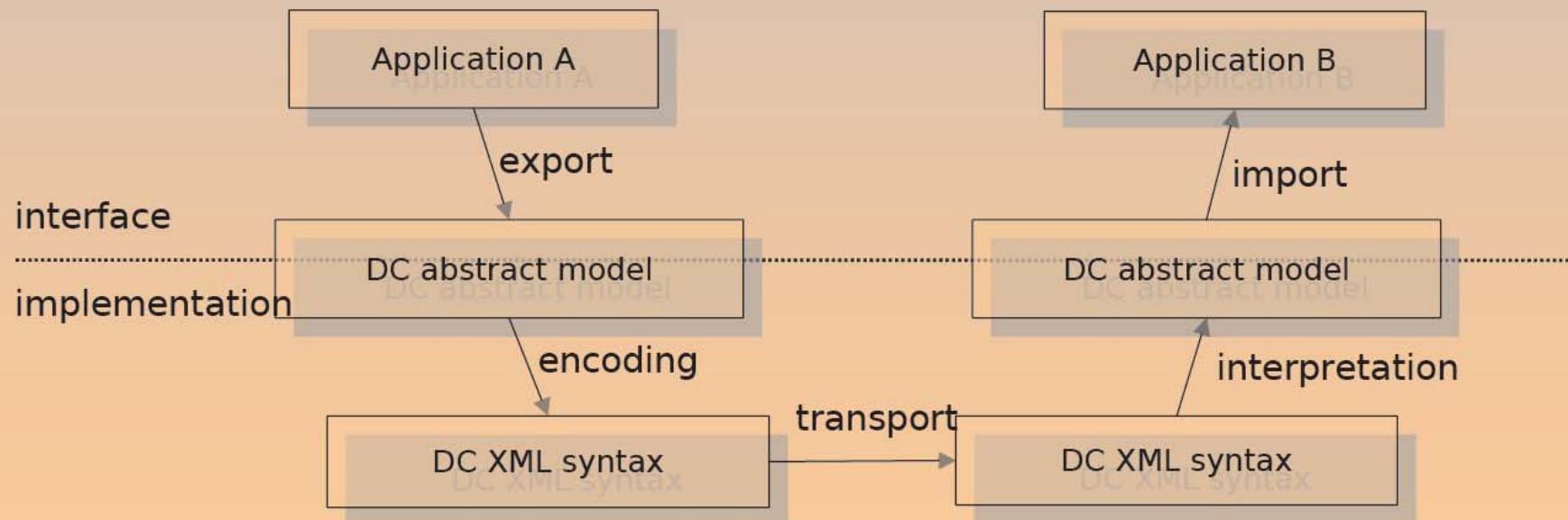


From DCAM to RDF to

- Real-world metadata applications tend to be based on loosely grouped sets of descriptions (where the described resources are typically related in some way)
- In the abstract model they are known as as description sets
 - for example, a description set might comprise descriptions of both a painting and the artist
- Description sets are instantiated, for the purposes of exchange between software applications, in the form of *metadata records*
- Each record conforms to one of the DCMI encoding guidelines (HTML meta tags, XML, RDF/XML, etc.)
- It is easy to express a DC description set as a RDF graph, and then express it in RDF/XML

Use of abstract model

- DCMI Abstract Model gives “interface”, or “conceptual model”
- syntaxes give realization/implementation



RDF – Resource Description Framework

- Resource Description Framework (RDF) is a language for representing information about *resources* in the WWW
- All *resources* are identified by a URI (URIrefs)
- *Resources* are described in terms of simple statements specifying properties and property values
- A statements is: (each component is indicated by a URIref)
 - A subject
 - A predicate (about the subject)
 - An object (the value of the predicate)

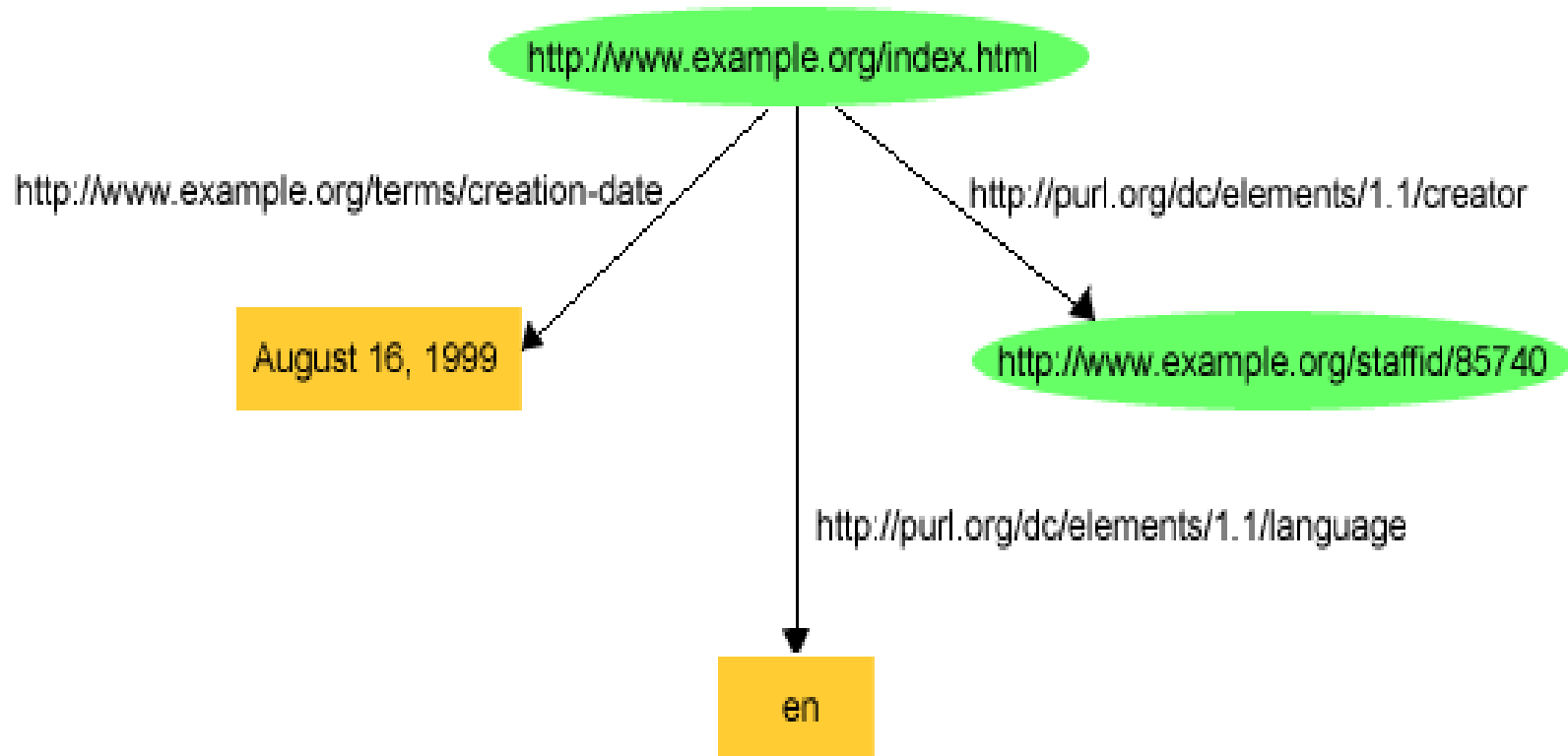
Simple RDF statement

- `http://www.example.org/index.html` "has a creator" whose value is John Smith
- the subject is the following URI
`http://www.example.org/index.html`
- the predicate is the word "creator"
- the object is the phrase "John Smith"
- To avoid "misunderstandings", all three indicated by URIs
 - Subject `http://www.example.org/index.html`
 - Predicate `http://purl.org/dc/elements/1.1/creator`
 - Object `http://www.example.org/staffid/85740`

Additional RDF statements

- <http://www.example.org/index.html>
has a creator
whose value is John Smith
- <http://www.example.org/index.html>
has a creation-date
whose value is August 16, 1999
- <http://www.example.org/index.html>
has a language
whose value is English

RDF is represented as a graph



RDF statements as triples

Each triple corresponds to an arc in the graph

- `<http://www.example.org/index.html>`
- `<http://purl.org/dc/elements/1.1/creator>`
- `<http://www.example.org/staffid/85740>` .

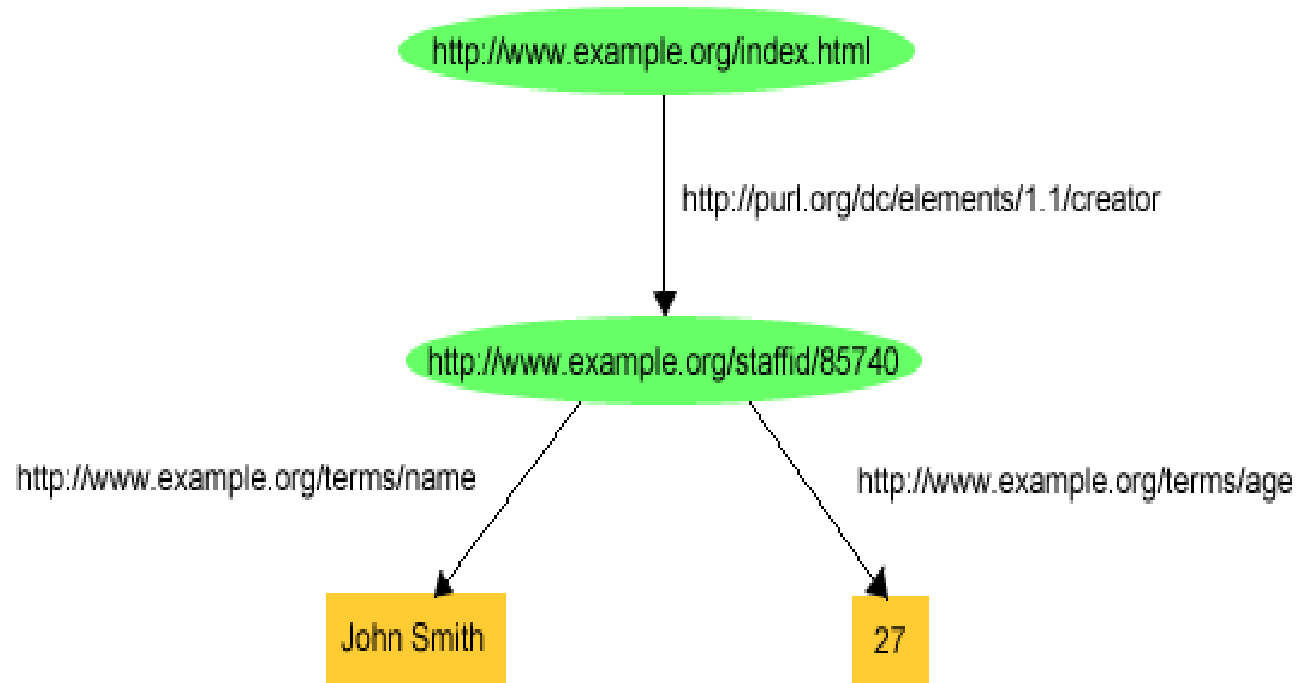
- `<http://www.example.org/index.html>`
- `<http://www.example.org/terms/creation-date>`
- `"August 16, 1999"` .

- `<http://www.example.org/index.html>`
- `<http://purl.org/dc/elements/1.1/language>`
- `"en"` .

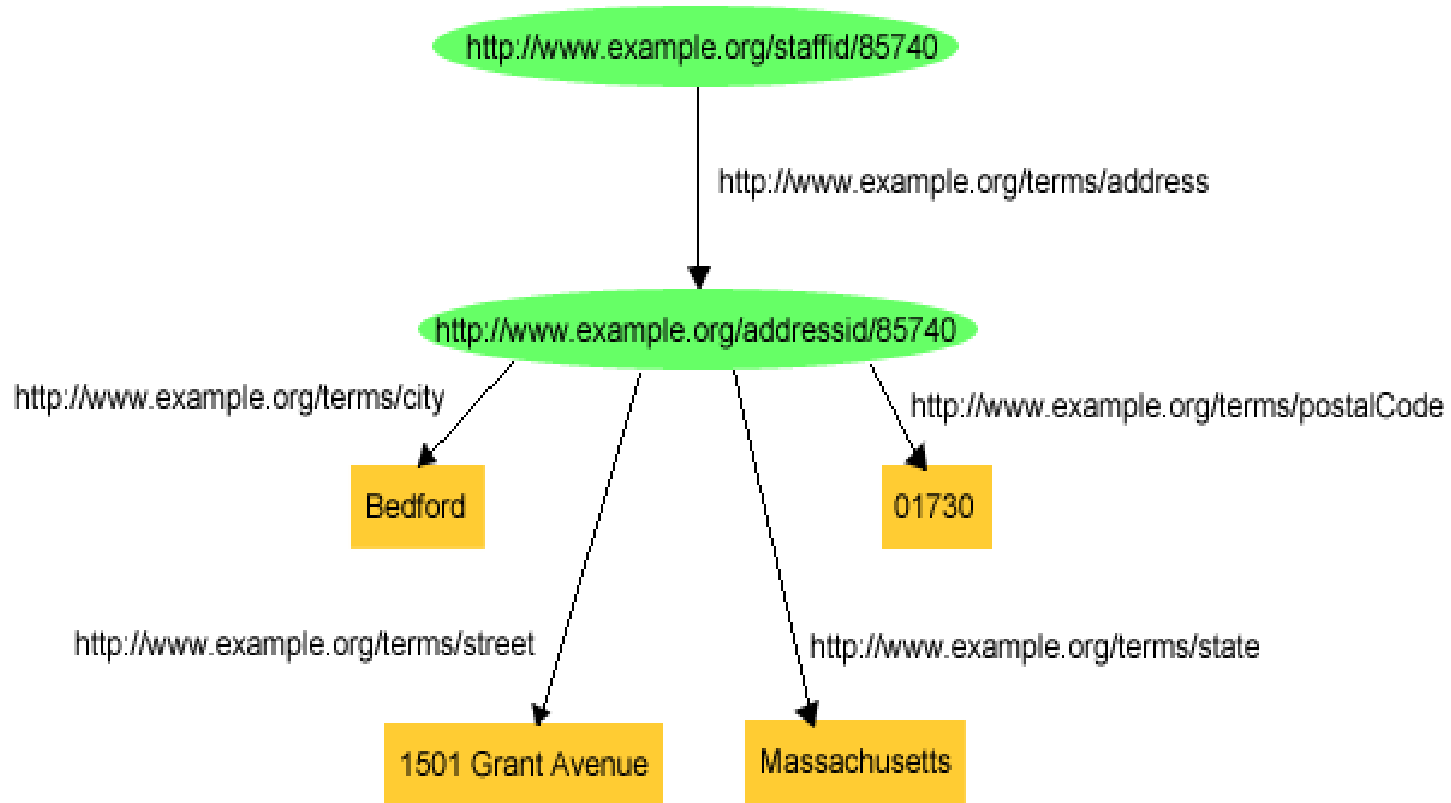
Abbreviated triples

- ex:index.html
dc:creator
exstaff:85740 .
- ex:index.html
exterms:creation-date
"August 16, 1999" .
- ex:index.html
dc:language
"en" .

URI values can be subjects ...



... to any depth ...



RDF summary

- An RDF graph is a set of RDF triples
- An RDF triple has three components
 - an RDF subject, which is an RDF URI reference or a blank RDF node
 - an RDF predicate, which is an RDF URI reference
 - an RDF object, which is an RDF URI reference, a blank RDF node or an RDF literal
- An RDF literal can be of two kinds
 - an RDF **plain literal** is a character string with an optional associated language tag describing the language of the character string
 - an RDF **typed literal** is a character string with an associated RDF datatype URI. An RDF datatype defines the syntax and semantics of a set of character strings that represent data such as booleans, integers, dates, etc.