

Metadata, FRBR, and New Challenges

Presentation for the
Missouri Library Association
St. Louis, Missouri, October 28, 2004

Presentation © 2004 Arlene G. Taylor

Remember When?

- Dictionary card catalogs
 - Prolific author
 - titles of works filed alphabetically behind guide card for author's name
 - each work might have a group of cards taking up an inch or more of space

Within a group of cards for a work you often found:

- original publication
- other editions, other formats
- translations (alphabetical by language)
- related works, e.g.:
 - movie based on book, songs from musical based on book, etc.
 - video versions
 - sound versions
 - text with commentary, etc.
- works in which this work is a subject (criticisms, reviews, etc.)

That was then, this is now.

- Today's online catalog at LC:
 - author search: Ferber, Edna
 - results: 106 entries
 - *Show boat* starts at entry number 79:
 - title alone (1929 publication)
 - microform version (1928)
 - title with subtitle "a novel" (1981 and 1926 publications)
 - title with statement of responsibility not subfielded (1943, 1926, and 1935 publications)
 - microform version (1974)
 - version containing 3 novels: *Show boat*, *So big*, and *Cimarron* (1962)

Today's LC online catalog:

- title search: Show boat
- results: 88 entries – default display:
 - sound recording of selections from Show boat (n.d.)
 - excerpts from Show boat and No, no, Nanette (1950)
 - excerpts from Show boat and Roberta (1960)
 - version of the movie (1989)
 - selections from Show boat and An American in Paris (1960)
 - versions of the movie (1951 and 1989)
 - Show stoppers in dance time [sound rec.] (1942)
 - Sound recording including "An incident from Show boat" (n.d.)
 - 5 of the 8 text versions shown under author Ferber

Today's LC online catalog (cont.)

- title search: Show boat display (cont.)
 - a "highlights" sound recording (1949)
 - themes sound recording (1942)
 - movie versions (1920 and n.d.)
 - sound recording of the movie with additions (1988)
 - sound recordings (1942 and 1946)
 - Theatre Royal Drury Lane production (1928)
 - movie version (1929)
 - a "highlights" sound recording (1957)
 - a play version (1934)
 - two more text versions (that were also found under Ferber - 1981 and 1926)
 - Show boat cookbook (1996)
 - etc., etc.

Today's LC online catalog (cont.)

- subject search: no hits
- ☞ OCLC search for author Ferber and Show boat in subject:
 - video of a parody of the movie and book (1978)
 - thesis on the work as literature and as film (1991)
 - journal review of the controversy around the revival of Show boat (analytical entry - 1995)
 - sound recording of Ferber reminiscing about Show boat (published 1983, although author died in 1968; so this must be a copy of an earlier recording)
 - Show boat cookbook (1996)
 - book about this "classic" musical (1977)
 - theses about the production of Show boat (1961, 1971, and 1985)

Not convinced yet?

- ☞ The examples on the last several slides did not include searches for "showboat" which retrieves different results from "show boat."

What Can Be Done?

- FRBR
- and, because FRBR was published in 1998 and therefore does not address virtual electronic resources except as downloadable copies of documents, I'll say a few words about the Semantic Web.

FRBR

- Functional Requirements for Bibliographic Records (IFLA, 1998)
 - What is it?
 - What does it attempt to accomplish?
 - Who is it for?
 - How is it being accepted and implemented?

What is FRBR?

- ☞ A conceptual model
- ☞ Based on the entity-attribute-relationship model of analysis
- ☞ Entities of interest to users of bibliographic systems are identified
- ☞ Attributes of interest to users are identified for each entity
- ☞ Relationships that operate between entities are specified.

Entities

- ☞ Three groups of entities
 - Group 1 – the products of intellectual or artistic endeavor that are named or described in bibliographic records (*work, expression, manifestation, and item*)
 - Group 2 – entities that are responsible for intellectual or artistic content, physical production and dissemination, or custodianship of such products (*person, and corporate body*)
 - Group 3 – entities in addition to those in groups 1 and 2 that may serve as the subjects of intellectual or artistic endeavor (*concept, object, event, and place*)

Group 1 entities

work

- distinct intellectual or artistic creation
- abstract entity with no single material object one can point to
- recognized through individual *expressions* of the work
- e.g., Ferber's *Show boat*; the musical, *Show boat*; Mozart's *The Magic Flute*; online journal *D-Lib Magazine*;

Group 1 entities (cont.)

expression

- realization of a work in alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc.; or any combination of such forms
- a new expression excludes aspects of physical form (e.g., typeface) that do not change intellectual or artistic realization of the work
- e.g., for Edna Ferber's *Show boat*
 - e_1 – 1926 ed. published by Country Life Press
 - e_2 – 1929 ed., illustrated with scenes from the movie
 - e_3 – version translated into Spanish
- e.g., for a work of Franz Schubert:
 - e_1 – the composer's score
 - e_2 – a performance by the Amadeus Quartet
 - e_3 – a performance by the Cleveland Quartet

Group 1 entities (cont.)

☞ manifestation

- physical embodiment of an expression of a work; when production involves changes in the physical form (or format), it results in a new manifestation
- e.g., Ferber's *Show boat*
 - e_1 – original text document
 - m_1 – Grosset & Dunlap (N.Y.) 1926 publication
 - m_2 – archival photocopy of m_1
 - m_3 – Heinemann (London) 1926 publication
- e.g., *The New York Times*
 - e_1 – paper (vs. Web) version
 - m_1 – print-on-paper format
 - m_2 – microfilm format

Group 1 entities (cont.)

☞ item

- a single exemplar of a manifestation
- normally the same as the manifestation itself
- however, variations external to the intent of the producer of the manifestation can occur in different items
 - e.g., damaged copy; copy autographed by author; copy bound by a library, etc.

Group 1 entities (cont.)

- ☞ Group 1 entities are the ones you've probably heard the most about and are the ones that are being used in models of "FRBRization"
- ☞ Top down vs. bottom up
 - bottom up easier to understand in the organizing process
 - also easier to catalog
 - if only one manifestation in hand, that's what needs to be described – how would you know what other manifestations and expressions exist or will exist, especially for a new item?
 - top down easier for users
 - makes more sense for display

Group 2 entities

- ☞ person
 - an individual, living or deceased
- ☞ corporate body
 - an organization or group of individuals and/or organizations acting as a unit
- ☞ a person and or a corporate body is a FRBR entity only if involved in the creation or realization of a *work*, or is the subject of a *work*

Group 3 entities

- Any group 1 or 2 entities may be the subject of any group 1 entity
- There are 4 additional group 3 entities:
 - concept
 - an abstract notion or idea – encompasses a comprehensive range of abstractions, and may be broad in nature or narrowly defined and precise

Group 3 entities (cont.)

- object
 - a material thing, including animate and inanimate objects occurring in nature; fixed, movable, and moving objects that are products of human creation; objects that no longer exist
- event
 - an action or occurrence (e.g., historical event, epoch, period of time)

Group 3 entities (cont.)

- place
 - a location, from one of a comprehensive range of locations: terrestrial and extra-terrestrial; historical and contemporary; geographic features and geo-political jurisdictions
- ☞ a concept, an object, an event, or a place is treated as a FRBR entity only if it is the subject of a work

Aggregate and component entities

- ☞ The model allows us to represent aggregate entities as one work, e.g.:
 - several works brought together by an editor in an anthology
 - monographs brought together by a publisher in a series
 - collection of private papers organized by an archive as a single fond
- ☞ or to have a component of a larger work (e.g., chapter; map segment; journal article) treated as a work itself
- ☞ Aggregates and component entities are treated as whole/part relationships

Attributes of entities

- ☞ serve as means for users to formulate queries and to interpret responses
- ☞ FRBR contains detailed lists of attributes for each entity and definitions of each attribute
- ☞ some attributes may appear to be the same as entities, but they really are different (e.g., the attribute "statement of responsibility" pertains to the labeling information, while the entity *person* is a representation of a creation relationship – even though the words for each may be identical: "Edna Ferber," perhaps)

Attributes in common

- ☞ It is instructive to observe some attributes that *work* and *expression* have in common:
 - title
 - form
 - date
 - other distinguishing characteristic
- ☞ However, these would not necessarily have identical character strings (e.g. *Show boat* for the work, but *Showboat* for an expression; *symphony* for form of work, but *musical notation* or *sound* for forms of expressions)

Attributes for Groups 2 and 3

- Attributes for Groups 2 and 3 are not specified in FRBR
- These are being worked on by an IFLA task group working on functional requirements for authority records

FRBR relationships

- In the model, relationships serve to show the link between one entity and another.
- The following 3 slides, reproduced from the FRBR document, show the high-level entity-relationship diagrams from FRBR that indicate at a generalized level how the entities are connected with one another.

Figure 3.1: Group 1 Entities and Primary Relationships

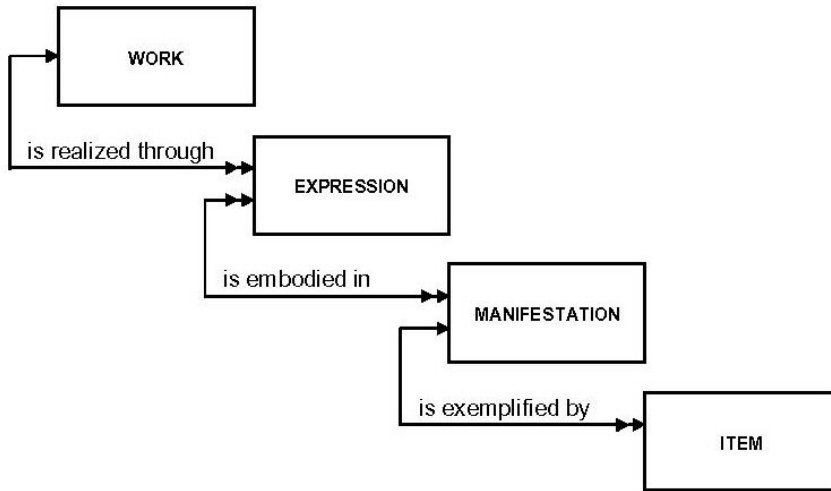


Figure 3.2: Group 2 Entities and "Responsibility" Relationships

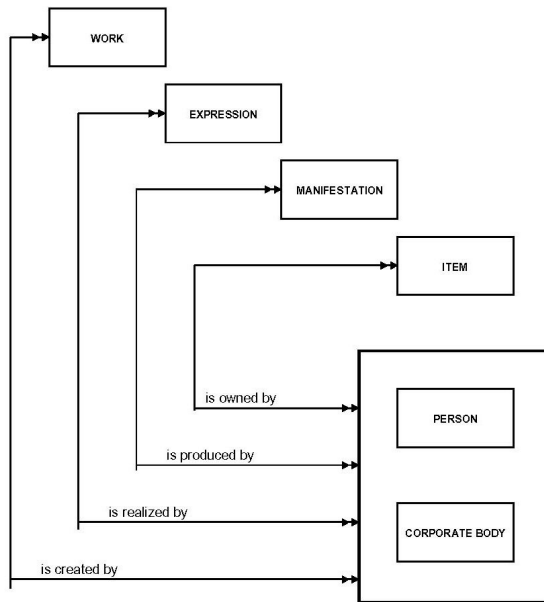
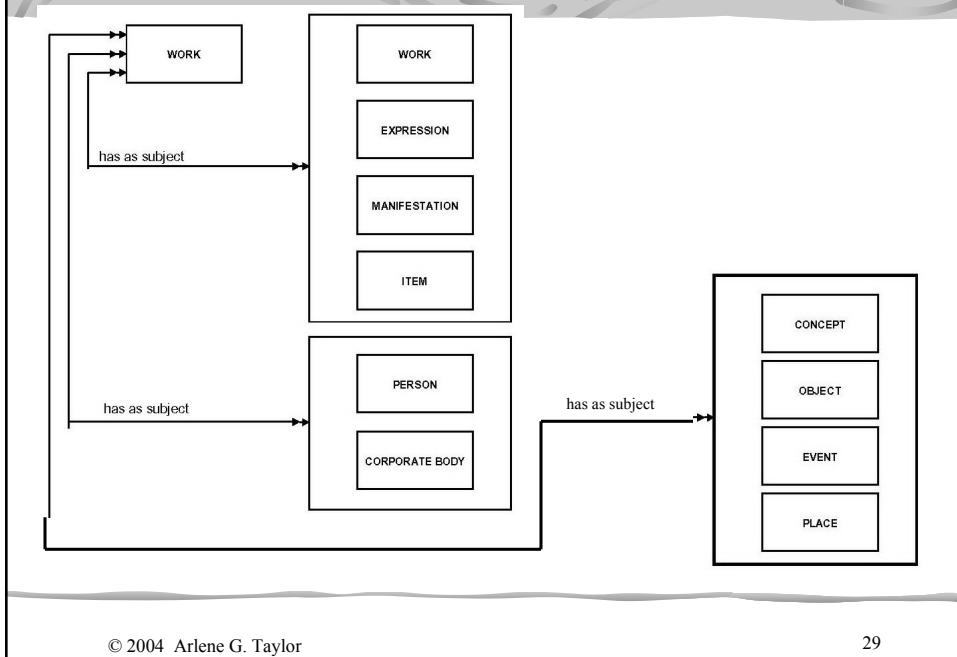


Figure 3.3: Group 3 Entities and “Subject” Relationships



Additional relationships

- ☞ In addition to these high-level relationships, FRBR identifies major types of relationships that operate
 - between instances of the same entity type
 - between instances of different entity types
- ☞ The only relationship type for manifestation-to-manifestation relationships, for example, is called “Reproduction,” and the kinds of reproduction listed are *reproduction*, *microreproduction*, *macroreproduction*, *reprint*, *photo-offset reprint*, and *facsimile*

Work-to-work

☞ Relationship types for work-to-work:

- successor (sequel, succeeding work)
- supplement (index, concordance, teacher's guide, gloss, supplement, appendix)
- complement (cadenza, libretto, choreography, ending for unfinished work, incidental music)
- summarization (digest, abstract)
- adaptation (adaptation, paraphrase, free translation, musical variation, harmonization, musical fantasy)
- transformation (dramatization, novelization, versification, screenplay)
- imitation (parody, imitation, travesty)
- whole/part (chapter, section, part, volume/issue, illustration for a text, sound aspect of a film, monograph in a series, journal article)

Expression-to-expression

☞ Relationship types for expressions:

- abridgement
- revision
- translation
- arrangement (music)
- successor [and the rest of the same list under "work-to-work relationships"]

☞ Barbara Tillett and Richard Smiraglia have done much work on bibliographic relationships. Tillett's diagram of Smiraglia's "bibliographic family" concept follows:

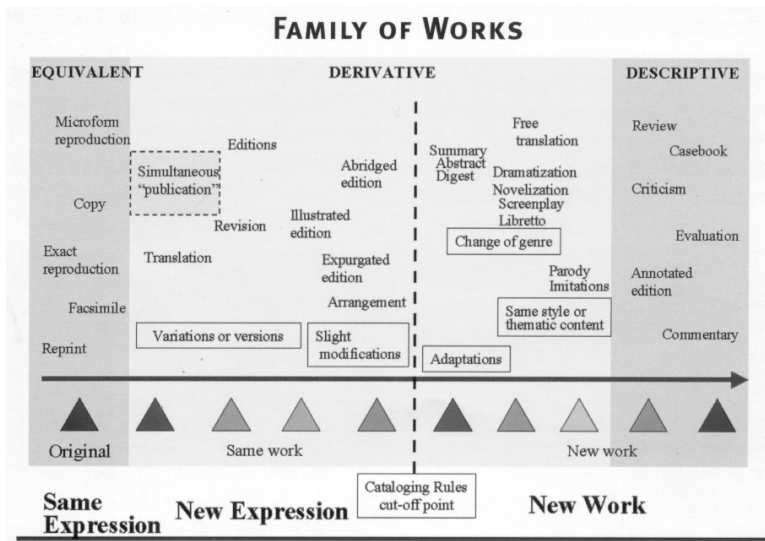


Figure from "Bibliographic Relationships" by Barbara B. Tillett (in *Relationships in the Organization of Knowledge*, ed. By Carol A. Bean and Rebecca Green, Boston: Kluwer Academic Publishers, 2001)

Other relationships

- ☞ Tables are also given for manifestation-to-manifestation relationships and for item-to-item relationships.
- ☞ See Functional Requirements for Bibliographic Records: Final Report, IFLA Study Group on the Functional Requirements for Bibliographic Records (Munich: Saur, 1998). Also available:
<http://www.ifla.org/VII/s13/frbr/frbr.pdf>

User tasks

- ☛ Finally, FRBR maps the attributes and relationships to user tasks:
 - to find entities that correspond to the user's stated search criteria
 - to identify an entity
 - to select an entity that is appropriate to the user's needs
 - to acquire or obtain access to the entity described

Practical applications?

- ☛ Several other countries have already moved to implement systems based on FRBR; the U.S. is somewhat slow here.
- ☛ AACR is being revised to explain rules in the context of entities, attributes, relationships, and user tasks.
 - probably will not change the creation of records at the manifestation level
- ☛ System design will be profoundly affected
 - OCLC and RLG are using the FRBR model to plan and design future systems
 - VTLS's Virtua system is modeled on FRBR

VTLS implementation

- ☞ Mixed – only 18-20% of catalogs need to be “FRBRized”
- ☞ Uses fixed field information and a few specified fields to create “work” level records and “expression” level records
- ☞ Uses a tree structure much like that in Windows Explorer to display the levels
 - clicking a + sign beside a work name displays the expressions
 - clicking a + sign beside an expression name displays the manifestations
 - items are in holdings records linked to manifestation records

VTLS FRBR Benefits

- ☞ According to Vinod Chachra and John Espley of VTLS, benefits are:
 - collocation
 - easier to find information
 - easier to see different expressions of a single work
 - easier to find all manifestations
 - better understanding of relationships between related works or expressions
 - in the circulation system a user can place “holds” at the work or expression level rather than only at the manifestation level as is currently required in most systems

OCLC's FictionFinder

- ☞ prototype system for 2.5 million records for fiction
- ☞ clustered using the OCLC FRBR Work-Set Algorithm
 - a work-set is a group of records that has the same author/title key
 - variant forms are brought into the work-set with the aid of the authority file
 - Alice in Wonderland, for example, presents 26,088 items as 1,267 manifestations in 51 languages
- ☞ FictionFinder demo: <http://fictionfinder.oclc.org/>

A Project of OCLC Research OCLC Online Computer Library Center Exit

FictionFinder

Search **Browse** **Categories**

Search **Clear**

Search for:

in:

AND

in:

AND

in:

AND

in:

Search **Clear**

(Basic Index: **Show boat**) AND Author: **Ferber**

1 4 Works were found.

1.	Divadelní loď (Show Boat), román	Vers:1	★	Ferber, Edna,
2.	Show Boat	Vers:43	★★★★★	Ferber, Edna,
3.	Show Boat, So Big, Cimarron: Three Living Novels of American Life	Vers:1	★★★★★	Ferber, Edna,
4.	Show Boat, So Big, Cimarron: Three Novels of American Life	Vers:1	★★★★★	Ferber, Edna,

1 4 Works were found.

[First work listed in this figure is "Show Boat" in Czech and really should be part of number 2.]

© 2004 Arlene G. Taylor 40

A Project of OCLC Research OCLC Online Computer Library Center Exit

FictionFinder

Search Browse Categories

Search

Search for:

in: All Fields

AND

in: Title

AND Ferber

in: Author

AND

in: Fiction Categories

Search Clear

(Basic Index: **Show boat**) AND Author: **Ferber**

(1) [Previous Record](#) Search Results (4 retrieved) [Next Record \(3\)](#)

Show boat, = Show boat

Work Set Key ferber, edna\1887 1968/show boat

Author [Ferber, Edna](#) G

Published Versions: 43, languages: 5

Fiction Categories [Domestic fiction](#)

[Historical fiction](#) E

[Love stories](#) E

Setting [Mississippi River](#)

Subjects [Frontier and pioneer life —Mississippi River Region](#) Q

[Gamblers](#)

[Interracial marriage](#)

[River boats](#)

[River life —Mississippi River](#) Q

[Showboats](#)

[Showboats —Mississippi River](#) Q

[Women entertainers](#)

Literary Forms [Fiction in English, 1900- Texts](#)

Special Forms [Large type books](#)

[English \(38\)](#)

[German \(2\)](#)

[Dutch \(1\)](#)

[Portuguese \(1\)](#)

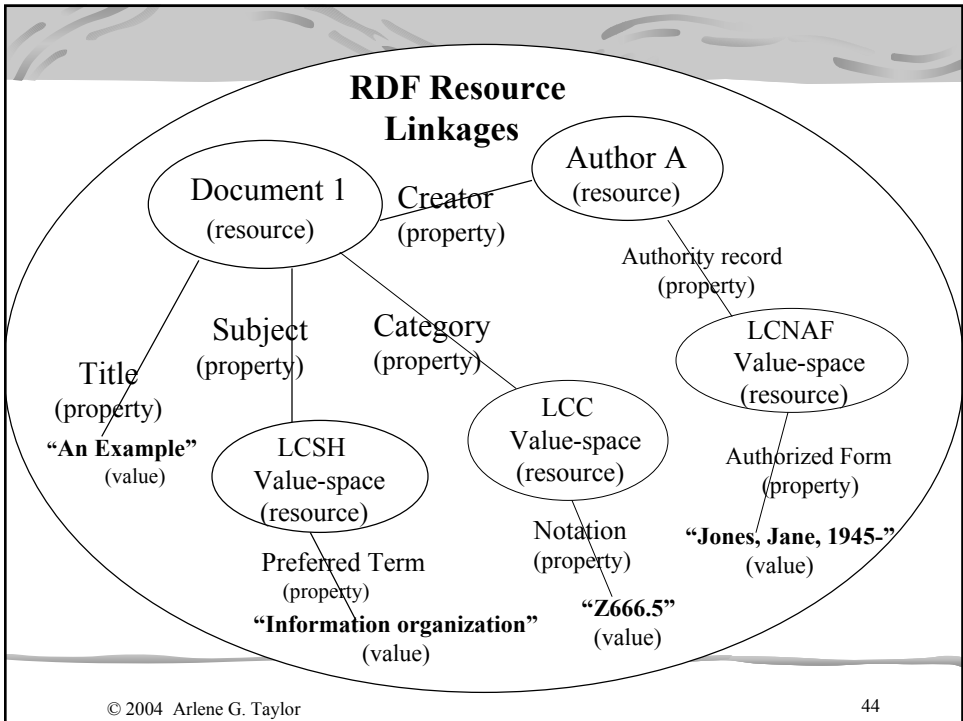
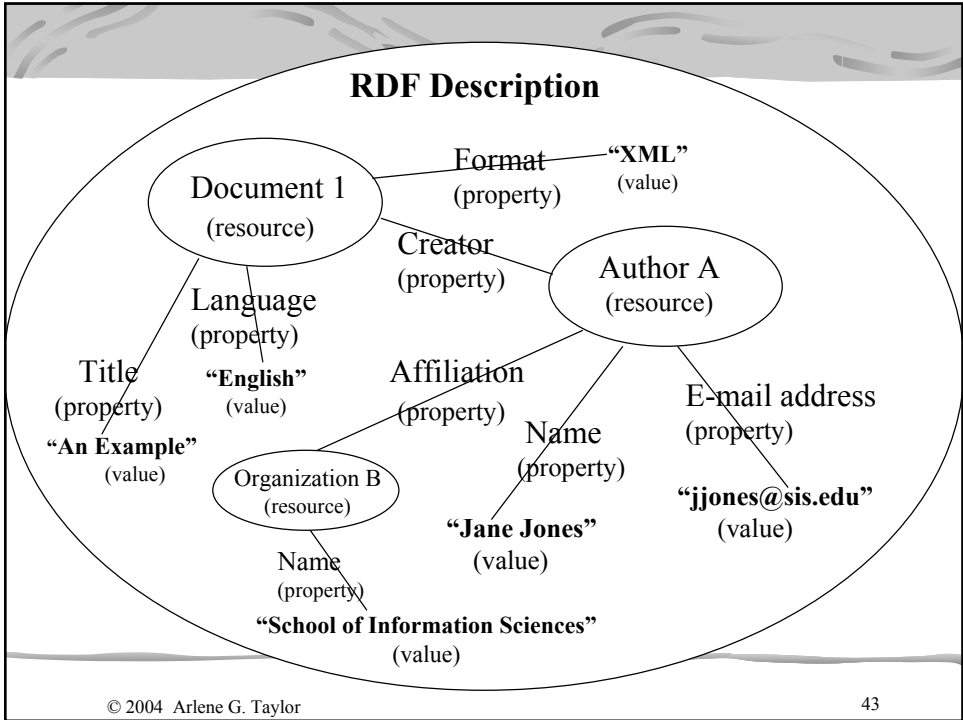
[Spanish \(1\)](#)

© 2004 Arlene G. Taylor 41

Models for Organizing Web Resources

- ☛ **RDF (Resource Description Framework)**
 - an infrastructure that enables the encoding, exchange, and reuse of structured metadata
 - uses XML as the means for exchanging and processing the metadata
 - based on the premise that resources have properties (or attributes), properties have values, some values can be other resources with their own properties and values, and all these relationships can be linked within the framework
 - amazingly similar to FRBR

© 2004 Arlene G. Taylor 42



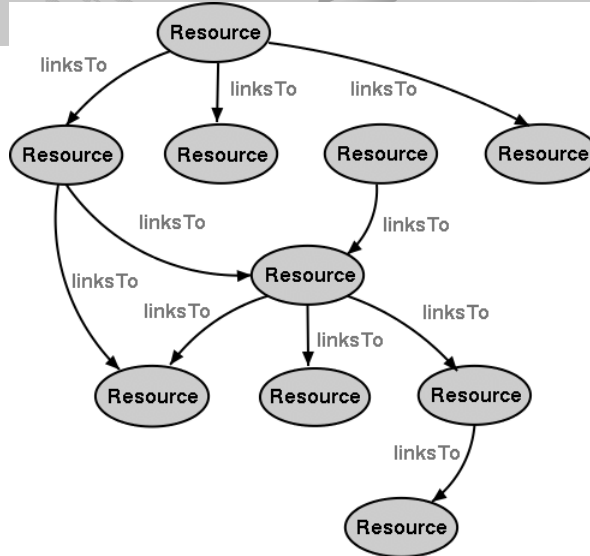
Semantic Web

- ☞ depends on "entity" metadata for resources (e.g., documents, people, corporate bodies, concepts, schemas, etc.)
- ☞ every resource has a unique identifier (e.g., URI) which can lead to the entity metadata that exists for the resource
 - value of the property of "creator" can be the URI for a person
 - this "access point," attached to every resource related to the same creator, fosters collocation
- ☞ it is intended that information will be defined in such a way that its "meaning" or "semantics" can be discernable, shared, and processed by automated tools as well as by people

Semantic Web (cont.)

- ☞ Berners-Lee, et al.: "For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning."
 - components required for this framework include URIs, XML, RDF, and ontologies
 - URIs provide a way to identify every resource as well as relationships among resources.
 - XML allows creators of Web pages to create their own tags to annotate sections of text or images – that is, to create structure for a document, if not meaning.
 - Meaning comes through RDF with its assertions that resources have properties (attributes) with values
 - Ontologies control values (e.g., authority files, controlled vocabularies, classifications)

Current Web



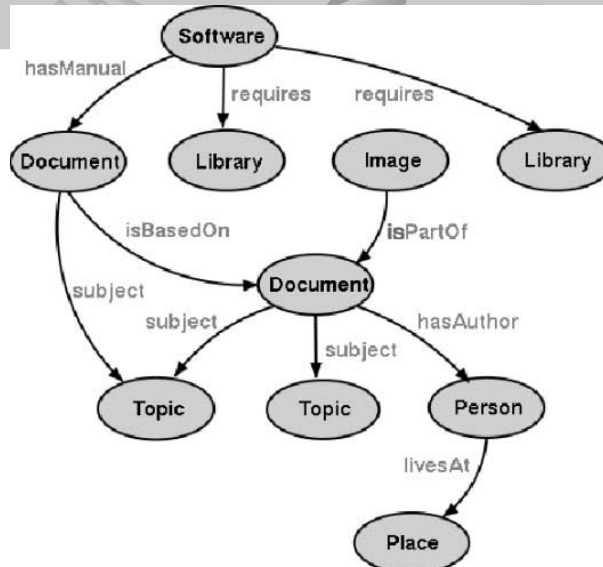
From Eric Miller's presentation, "Weaving Meaning: An Overview of The Semantic Web" at the Semantic Web Website:

<http://www.w3.org/2004/Talks/0120-semweb-umich/slide5-0.html>

© 2004 Arlene G. Taylor

47

Semantic Web




From Eric Miller's presentation, "Weaving Meaning: An Overview of The Semantic Web" at the Semantic Web Website:

<http://www.w3.org/2004/Talks/0120-semweb-umich/slide6-0.html>

© 2004 Arlene G. Taylor

48



Thank you.

Further questions may be addressed to:

ataylor@mail.sis.pitt.edu