Linked Data and the Semantic Web

Overview for BIBFRAME Pilot 2.0 Participants

Context

- AACR2, RDA, MARC 21 record environment
- Library records live in silos of data
- Not integrated with the web
- Linked data techniques promise a possibility to increase the visibility and usage of library data on the Web

Linked data defined

- A set of best practices for publishing and connecting structured data on the Web http://linkeddata.org/faq
- Key technologies support linked data:
  - URIs/IRIs that identify resources and entities
  - HTTP as a mechanism to retrieve data
  - RDF model to structure and link data

The Semantic Web

- "*The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.*"

What enables a Web of Data?

- Use a set of best practices for publishing and linking structured data on the Web.
- Use technologies that are more generic, more flexible which make it easier for data consumers to discover and integrate data from large number of data sources and links

Traditional Web vs. Semantic Web

Web of Documents

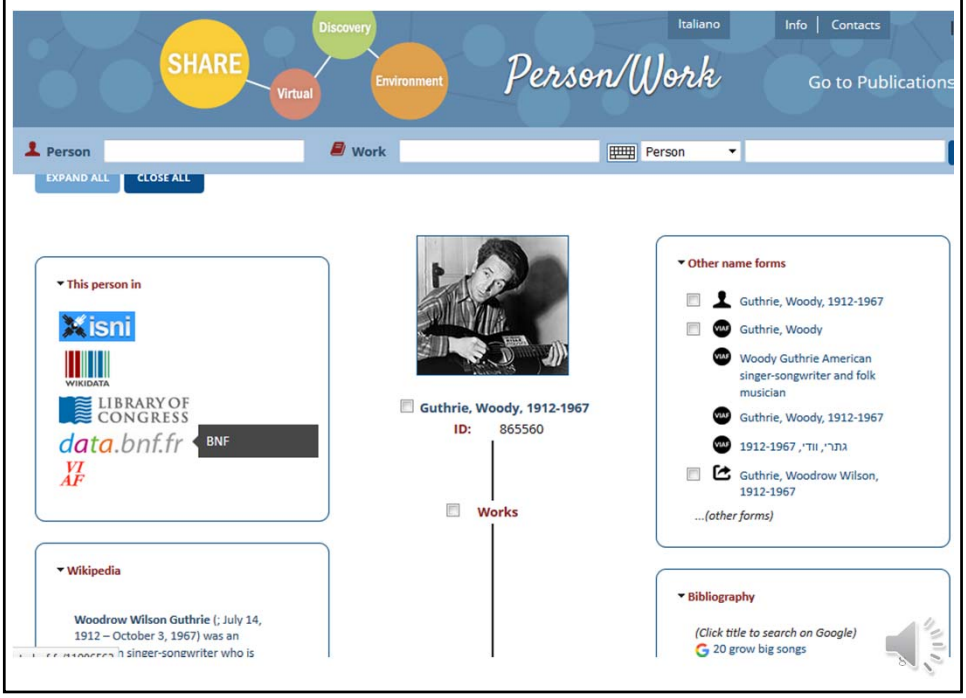
- information resources
- links between documents
- unstructured data
- implicit semantics
- for human consumption

Web of Data

- "real-world objects"
- links between things
- structured data
- explicit semantics
- for machines and human

Why linked data?

- Increase the visibility and usage of Library data on the Web
- Integrate library data with the large number of structured data sources and links on the web
- State relationships among resources
- Enhance the sharing of library data with a wider audience
- Facilitate a fuller implementation of RDA



Experimentation with BIBFRAME

- Linked data as a carrier of library data
  - a standard machine readable format
  - using common web standards
- Transition: from a static two-dimensional collocated record to decentralized data with links to illuminate relationships

MARC Record

WorldCat: Women at the Hague

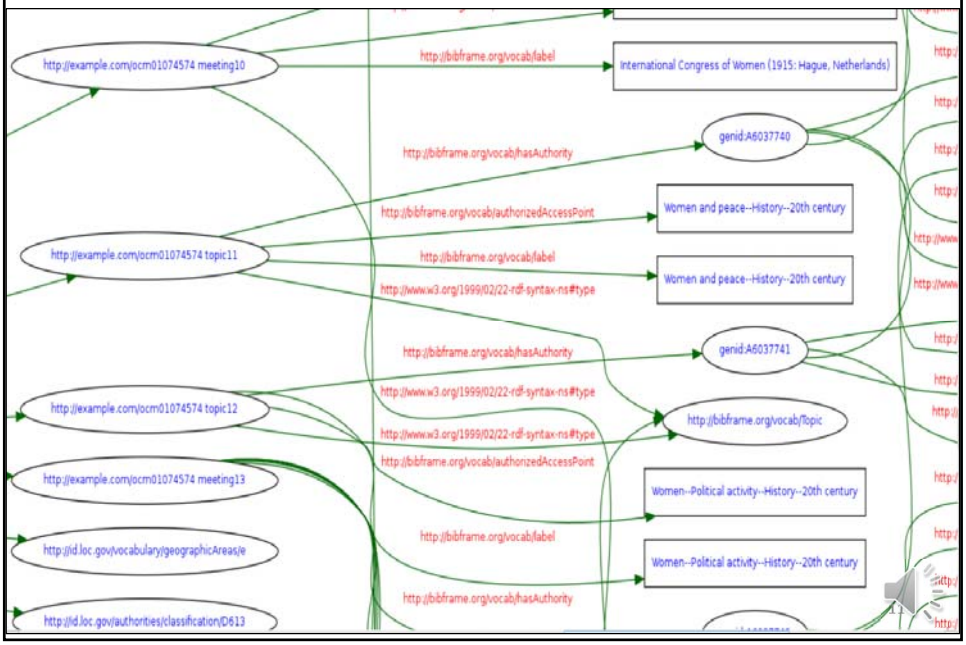
OCLC 1074574 — CLU Holdings - 294 other holdings; 10 other IRs
1 CLU holding in GLIMIR cluster; 295 other holdings in GLIMIR cluster

Books	Rec Stat	c	Entered	19741113	Replaced	20140821141935.9			
Type	a	ELvl	I	Srce	Audn	Ctrl	Lang	eng	
BLvl	m	Form		Conf 0	Biog	MRec	Ctry	nyu	
		Cont		GPub	LitF 0	Indx 0			
Desc	i	Ills		Fest 0	DtSt s	Dates 1915,			

010		15025357
040		DLC $b eng $c TLM $d CRL $d OCLCQ $d OCLCG $d STF $d NIALS $d AMAZN $d UKMGB $d BMC $d OCLCA $d MIGCL $d OCLCQ $d OCLCO $d HIN $d VA@ $d DEBBG
016	7_	000017737 $2 Uk
016	7_	012785073 $2 Uk
019		123238394
043		e------
050	00	D613 $b .A4 1915
082	_4	940.3/12
084		MS 3150 $2 rvk
084		NW 8100 $2 rvk
049		CLUM
100	1_	Addams, Jane, $d 1860-1935.
245	10	Women at the Hague: $b the International Congress of Women and its results / $c by three delegates to the congress from the United States, Jane Addams, Emily G. Balch ... Alice Hamilton ...
260		New York : $b The Macmillan Co., $c 1915.
300		vii, 171 pages ; $c 18 cm
336		text $b txt $2 rdacontent
337		unmediated $b n $2 rdamedia
338		volume $b nc $2 rdacarrier
505	0_	Journey and impressions of the Congress / Emily G. Balch -- At the war capitals / Alice Hamilton -- The revolt against war / Jane Addams -- Factors in continuing the war / Jane Addams -- At the northern capitals / Emily G. Balch -- The time for making peace / Emily G. Balch -- Women and internationalism / Jane Addams.

Humans can 'connect the dots,' but …

RDF Graph of a MARC Record



Tim Berners-Lee describes the Semantic Web

Tim Berners-Lee, James Hendler and Ora Lassila, "The Semantic Web", *Scientific American*, May 2001, p. 29-37.

RDF: Resource Description Framework

- Standard model for exchange of data on the Web
- Structures relationships between resources, people, and things on the web
- Uses graph model to represent database relationships
- RDF and related standards maintained by the World Wide Web Consortium (W3C)

RDF data model

- **Triple statements**
- **URIs and IRIs**
- **Ontologies and vocabularies**
- Graph data model
- RDF XML (or other serialization format)
- Namespaces

URIs and IRIs

- URI: uniform resource identifier
  - Sequence of characters used to identify a resource
- IRI: internationalized resource identifier
  - Identifier with extended character set
- This presentation uses the term “URI” for both of these concepts

URIs on the semantic web

- On the web of documents URL is a type of URI that links documents
- On the semantic web, URIs identify real-world objects
  - People
  - Cars
  - Books
  - Unicorns

URIs in RDF

- URIs identify resources
  - Such as a book or author
  - Namespaces of standards that have been used to encode RDF triple statements
  - Vocabulary and ontology terms
  - Subject, predicate, and object in triple statements

Triple Statements

- **Subject:** identifies a “Resource of interest”
- **Predicate:** identifies a Property of the “resource of interest” -- a relationship
- **Object:** identifies a Property value -- a resource that has a relationship to the “resource of interest”

Triple statements

Subject → Predicate → Object

This work → Was written by → This author

Triple statements

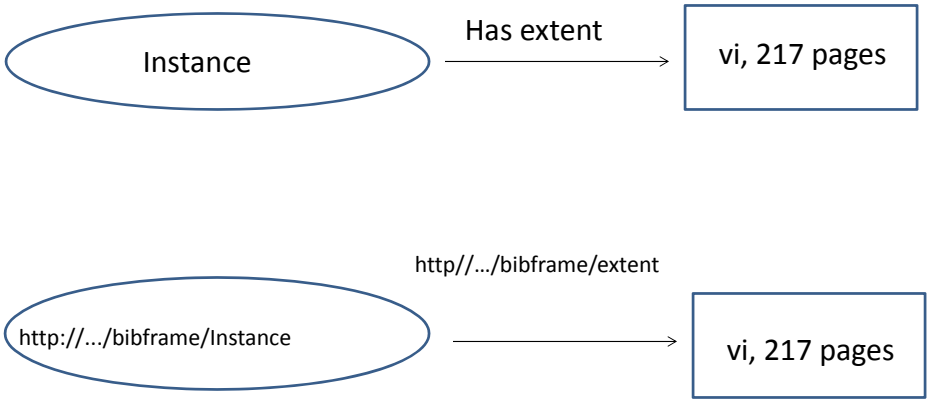
This land is your land → Was written by → Woody Guthrie

URI for work → URI for term: Composer → URI for author

Literals

- Non-URI values
- Used to identify values such as a strings, numbers and dates.
- Literals may only appear in the **object position** of a triple.

Example: Literal



Blank nodes

- A blank node is a local identifier rather than a global URI.
- They can be useful when you need to link to a collection of items.
- Blank nodes can appear in the subject and object position of a triple

Vocabularies and Ontologies

- Used to define concepts within a particular field of study (domain)
- Classify terms used in a particular domain
- Used to state relationships between resources
- Are necessary for discovering relationships on the Semantic Web

Current state of Linked Data in libraries

- Developing use cases
- Structuring, cleaning and releasing data
- Developing new frameworks and tools
- Exploration, prototypes and proofs of concept
- Learning!