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

<table>
<thead>
<tr>
<th>Web of Documents</th>
<th>Web of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>information resources</td>
<td>“real-world objects”</td>
</tr>
<tr>
<td>links between documents</td>
<td>links between things</td>
</tr>
<tr>
<td>unstructured data</td>
<td>structured data</td>
</tr>
<tr>
<td>implicit semantics</td>
<td>explicit semantics</td>
</tr>
<tr>
<td>for human consumption</td>
<td>for machines and human</td>
</tr>
</tbody>
</table>
**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

Humans can ‘connect the dots,’ but ...
RDF Graph of a MARC Record

Tim Berners-Lee describes the Semantic Web

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**: This work
- **Predicate**: Was written by
- **Object**: This author

This work is written by this author.

---

### Triple statements

- **Subject**: This land is your land
- **Predicate**: Was written by
- **Object**: Woody Guthrie

This land is your land is written by Woody Guthrie.

- **URI for work**: URI for term: Composer
- **URI for author**: URI for author

URI for work and URI for term: Composer are connected to 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

Instance ➔ Has extent ➔ vi, 217 pages

http://…/bibframe/Instance ➔ http//…/bibframe/extent ➔ vi, 217 pages
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!