Introduction to Linked Data

Lesson 1 Transcript

Hello, and welcome to the BIBFRAME video tutorial series, presented by the Cataloging Policy Specialists in the Policy, Training, and Cooperative Programs Division at the Library of Congress. The scope and purpose of this series is to provide training on using BIBFRAME.

This video tutorial is Lesson 1, Introduction to Linked Data.

By the end of this video tutorial, participants will be able to identify the basic concepts behind the Semantic Web, Linked Data, and RDF. Participants will also understand how linked data can benefit libraries and how linked data can affect their work.

The definitions of the Semantic Web and Linked Data have been evolving over the last twenty-five years, and any articles and videos you encounter might provide short but incomprehensible definitions, long and convoluted ones, or even definitions that conflate the two terms. For today's video tutorial, I will rely on a well-researched definition found in a 2019 article in the Journal of the Association for Information Science & Technology. In this article, the authors trace how concepts like the Semantic Web and Linked Data have changed over the years.

Specifically, the evolving conceptualization of the terms follows the movement of the Semantic Web from abstract to concrete, as operational developments in Linked Data bring the goals of the Semantic Web closer to reality.

The authors conclude that the Semantic Web is "a comprehensive construct that includes three components": one, the goal of making the web into a global database; two, the specifications that are responsible for a web created by the connection of data using open standards; and three, the set of rules and procedures for publishing data that can be connected to other data.

Notice how Linked Data is inextricably tied to the Semantic Web. Linked data is a set of rules for publishing data that can be reused and shared, consequently expanding the Semantic Web.

Tim Berners-Lee, the inventor of the World Wide Web, coined these four principles of linked data design in 2006. First, use unique identifiers, specifically Uniform Resource Identifiers or URIs, to name or identify things, from real world objects to abstract concepts. Second, allow these URIs to be looked up using the Hypertext Transfer Protocol, or HTTP. Third, when someone looks up a URI, they should be provided with useful information. And finally, in the information that is provided, give links to other URIs so more things can be discovered.

Linked Data allows us to state, in a machine-understandable way, that some piece of data has a particular relationship with another piece of data. It leverages standard web technologies -- like URIs and HTTP -- to make connections between data and to make them available on the web.

The relationships between pieces of data are indispensable to linked data and the semantic web, and these relationships are established and expressed through the Resource Description Framework (RDF)

data model. RDF is a model for describing resources and the relationships between them. Its basic structure is a triple comprised of a subject, an object, and a predicate.

The subject is a resource, and resources can be anything, including documents, people, real world objects, and abstract concepts. The subject is represented by an IRI, or an Internationalized Resource Identifier. The predicate is also represented by an IRI, and it states the relationship between the subject and object. The relationship is phrased in a directional way, from the subject to the object. The object can be an IRI or a literal, and it is the value represented by the property in the predicate.

These are some informal triple statements. We have two entities -- Carol and Moo Deng -- that are subjects in six statements. Moo Deng is also an object in one of the statements. A human being, with a basic understanding of how grammar works, can read these sentences and understand that Carol is a person who lives in Hawaii and is interested in a pygmy hippo named Moo Deng. Meanwhile, a computer "reads" each statement separately and then makes inferences based on logical connections.

These triples can be visualized as a connected graph consisting of nodes -- the subject and objects -- and arcs -- the predicates. Once you have a graph like this you can use a query language like SPARQL to look for data like people who are interested in Moo Deng.

So how does this relate to your work at the Library, unless you're interested in forming a hippo fan club?

Many of the changes brought by the Library's adoption of linked data practices will be absorbed by the automated systems that we use. For example, cataloging displays will be enriched with data drawn from the wider web to present to users. A linked-data friendly online catalog might include enhanced search results with author cards or panels that function like Google knowledge graphs. This author panel comes from the University of Miami Libraries. It collects data from Wikipedia, Wikidata, and the Library of Congress Name Authority File. It also includes links to library books by and about Augustine of Hippo and links to people related to him.

With linked data, libraries can continue to fulfill their responsibility to be sources of open and authoritative data. At the same time, moving to linked data gives library staff the opportunity to rethink how we consider library data; to change policies and practices; and to demonstrate our openness to learning new ontologies, new systems, and new content standards. All of these changes won't occur overnight. We will be working in a hybrid environment, using the BIBFRAME linked data editor to create metadata descriptions and then converting them into MARC bibliographic records that include linked-data friendly fields and subfields.

Thank you for watching this video tutorial. You can find more training, information, and support online at the Library of Congress.