

TESTIMONY

Linked Data: A Brief Introduction for Catalogers



Jonathan Edwards

Preacher

Jonathan Edwards was a Protestant preacher, philosopher, and theologian. Edwards "is widely acknowledged to be America's most important and original philosophical theologian," and one of America's greatest intellectuals. [Wikipedia](#)

Born: October 5, 1703, East Windsor

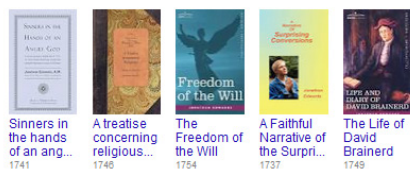
Died: March 22, 1758, Princeton, NJ

Spouse: Sarah Pierpont (m. 1727)

Buried: Princeton Cemetery, Princeton, NJ

Children: Esther Edwards Burr

Books



Try a Google search for Jonathan Edwards, and to the side you'll see information about his life and family, along with a display of books he wrote. If you click on one of the books, you may see a brief summary, a rating from Amazon or Goodreads, or links to web stores offering the book for sale, but as of this writing you won't see links to local libraries owning a copy. Why? Because our catalog records are still records, in MARC, two things that search engines such as Google and Bing can't understand. Search engines run on data, not records, and with more and more people beginning their research on Google, libraries need to make their catalog data findable on the Internet, or their catalogs will become obsolete.

Enter Linked Data.

That Google sidebar (called a knowledge card) doesn't come from an individual record that somebody created but instead pulls together pieces of data (such as personal names, birth dates and place names) that are linked together in ways that tells computers how each bit of data is related to another, so that a search engine knows that Jonathan Edwards isn't a text string but a person, a man born in East Windsor and buried in Princeton Cemetery. This is the general idea of Linked Data: bits of machine-readable data linked together to create meaning.

Here's how it works. Bits of data are identified by uniform resource identifiers (URIs) instead of text strings. This means, instead of identifying Jonathan Edwards with the text string **Edwards, Jonathan, 1703-1758**, as we currently do in a MARC 100 or 600 field, we would identify him with the URI <http://id.loc.gov/authorities/names/n79084179.html>, which is his unique identifier in the Linked Data version of the LC Name Authority File. This URI is then linked to the label **Edwards, Jonathan, 1703-1758**, marking it as LC's preferred label for Jonathan Edwards. In the Linked Data environment, that label can change multiple times, but the URI will remain the same. If somebody discovers that Edwards was born in 1702 instead of 1703, the label can be centrally updated to reflect that change, and library catalogs linking to that URI would receive an automatically updated label with no manual updating required.

Linked Data can be difficult to imagine in the abstract, so a great way to see it in action is via the [Reasonator](#), a tool that visualizes data from Wikidata (Wikipedia's Linked Data knowledge base). Try playing around with the page for [Jonathan Edwards](#), which currently displays birth and death dates, links to people that influenced his thought, and other data. You can find other people born in East Windsor, Conn., or others who died of the smallpox like Edwards did. These pages aren't crafted by hand, like individual catalog records, but are instead collocations of data pulled from various places in Wikidata. Nobody curated this data in advance, but Linked Data's flexibility allows for data curation on the fly, for applications that nobody had to think of in advance.

Thinking of cataloging in terms of linked data requires a paradigm shift. Put simply, catalogers are still creating digital catalog cards. We need to get past the idea of creating records and think in terms of creating pieces of data and linking them together in ways that can form "records" such as the ones in the Reasonator. This is where [BIBFRAME](#), the possible replacement for MARC, comes in. It remodels our catalog data, breaking up those digital catalog cards into pieces of data. While a full description of BIBFRAME is outside the scope of this article, a broad visualization of how it separates the data can help it seem less abstract. Below is a MARC record segment describing *A Treatise Concerning Religious Affections*. This is how we currently understand a description for a book: a page of data about a particular book.

100 10 \$a Edwards, Jonathan, \$d1703-1758
 245 12 \$a A Treatise concerning religious affections / \$c by Jonathan Edwards.
 260 0 \$a Glasgow : \$b Chalmers and Collins, \$c 1825.

Figure 2 shows the same data separated into pieces according to a simplified version of the Resource Description Framework (RDF), the data model underlying BIBFRAME. Some of the data chunks have URIs with them, providing them with unique identifiers that give them the potential to link to other data about the same entities. The URI for Jonathan Edwards links to his Virtual International Authority File (VIAF) page which links to a Wikidata page that contains additional data. The Glasgow URI links to the corresponding page on GeoNames, which states that Glasgow is a city in Scotland, with particular geographic coordinates. These external links not only make it possible to draw in supplementary data, but they also allow non-library sources to draw in our data, increasing visibility and usefulness.

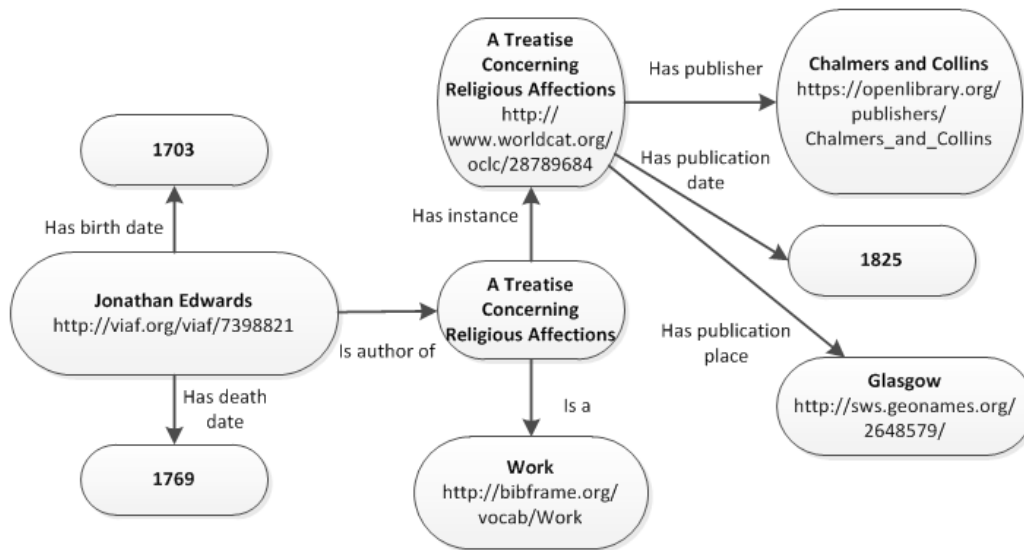


FIGURE 2: SIMPLIFIED RDF VISUALIZATION OF A TREATISE CONCERNING RELIGIOUS AFFECTIONS

Now what does this have to do with Google? BIBFRAME's RDF data model is a web standard that search engines can read. Those Google knowledge cards come from RDF data. Bibliographic information encoded in RDF plays nicely with web standards in a way that MARC records never will. As part of the BIBFRAME project, the team at Zepheira (the organization developing BIBFRAME) asked employees of Microsoft and Google why library resources don't appear in search engine results. The answer? They can't see us. They can't access our data, they can't read it, they can't harvest it.¹ To bridge this gap, a new project called the [Libhub Initiative](#) is gathering MARC records from participating libraries and converting them to BIBFRAME, then publishing them on the web to monitor their search engine visibility. Although still in its early stages, the project is already showing positive results for the early adopters.²

In addition to providing greater web visibility, catalog data as Linked Data offers other possibilities for use. A special collection of missions materials could be searchable via a map interface linked to the places described in the collection. A catalog linking bibliographic data to author data could allow a student to limit a search to commentaries written by Eastern Orthodox authors, without catalogers needing to specify an author's religious affiliation in a subject heading.

¹ Jeffrey Penka, *Linked Data and Why It Matters in the Library Community*. Vimeo video, 51:41, from a Zepheira webinar on June 2, 2015. <https://vimeo.com/130109172>.

² "Libhub Initiative Celebrates Founding 12 Library Partners at Early Adopter Summit," Zepheira.com, last modified May 18, 2015, <http://zepheira.com/news/pr20150527libhub-ea-summit/>.

A scholar could search for items written by Jonathan Edwards and others associated with the First Great Awakening without having to search for each individual separately. An authority record could automatically update without a cataloger needing to change a heading for every death date discovered.

Library Linked Data won't become widespread reality tomorrow. Experimentation is still ongoing. But, if we want our bibliographic descriptions to be more findable, more flexible, it's an idea we should embrace. We need to start putting our data where our users can find it.

For Further Learning

Linked Data for Libraries. YouTube video, 14:13, posted by "OCLCVideo" on August 9, 2012.

<https://youtu.be/fWfEYcnk8Z8>.

Linked Open Data - What is it? YouTube video, 3:42, posted by "EuropeanaEU's channel" on February 28, 2012.

<https://youtu.be/uju4wT9uBIA>.

Mitchell, Erik T., *Library Linked Data: Research and Adoption*. Chicago: ALATechSource, 2013.

Sporny, Manu, *What is Linked Data?* YouTube video, 12:09, posted on June 16, 2012.

https://youtu.be/4x_xzT5eF5Q.

Submitted by
Christa Strickler, Assistant Professor of Library Science
Buswell Memorial Library, Wheaton College