

Searchable Bibliographic Data

Anna Arenas

University of South Florida

The web originally contained information targeted mainly for “human consumption”. In other words, users were able to find information by following taxonomies (directories, as they were called) or by searches which easily provided results, not always accurate, but definitely fast since they were targeting pages. The web as it was known then comprised only links to pages. As content grew exponentially in quantity, search engines worked on ways to produce relevant results using not only the content and keywords embedded in the web pages but others factors as well, which included backlinks and popularity rankings among others.

Linking pages made the process of searching so easy, that it quickly became the common way to find information. Library old card catalogs became obsolete and libraries found themselves looking for new ways to offer accessible resources to their users. At that point, libraries targeted to make their content accessible through the web.

In time, the web transformed to semantic web. Pages aren't the ones linked, but the pieces of information they contain. Search engines like Google adopted schema.org - a structured data markup- developed specifically for web purposes, which allowed machine understanding of the content of web pages and consequently allowing better rankings for relevant search results. Schema.org was, and continues to be, intended for search engine purposes and it is supervised by web developing companies like Google, Yahoo, Microsoft, and more. Here is perhaps, where the main difficulty resides. The original purpose of schema was to index the content of web pages and allow the interconnection of information through the web, but it wasn't originally intended for bibliographic records. This distinction is key, because bibliographic records require content that might be irrelevant for most web users but are

fundamental to keep the full integrity each unit of information. It is fundamental to preserve all aspects of the data. On one hand, we have search engines that need to have a way to link data in order to provide the fastest and most accurate set of results to their users. Conversely, libraries want to open their content to the web and make it accessible while, at the same time, have in place a complete descriptive set of metadata that accompanies it.

The web is a decentralized structure and schema.org manages to link information items organically. Decentralization allows data to be “disentangled from local database systems and published in web-compatible formats” (Seeman, 332) Because it is decentralized, it also has the possibility to improve in its own. New indexing pushes the creation of new kinds of structured data. Nonetheless, the growth of schema.org doesn’t seem to be enough -in quality and speed- to what bibliographic records need.

Librarians have centuries of experience classifying and cataloging materials. They have always “been focused on making connections, whether in the construction of a cross reference or, now, in the application of a URI used to interconnect data”(DeWeese, 25) There is one premise that stands out in cataloging by librarians, and that is the ability to organize by disciplines and expose relationships between entities. Allowing librarians to organize information materials for the web can improve notably the search results.

Now the question comes if it is viable to improve schema.org to include all the metadata information and library workflows that bibliographic -and for that matter, objects, sound recordings, and visuals- require or is it a better solution to create a new system with a “vocabulary for libraries to express their collections on the web [...] consistent with Semantic Web best practices.” (Fons, 21)

For that matter, the Library of Congress has embarked in the development and application of BIBFRAME (Bibliographic Framework Initiative) a special vocabulary for bibliographic records that is consistent with current languages like schema.org. BIBFRAME makes “library records to conform to web standards” (El-Sherbini, 78) BIBFRAME ontology is built using RDF (Resource Description Framework) in such a way that it is interoperable with outside ontologies and it allows metadata to exist in a linked environment. “BIBFRAME was developed to be content standard agnostic” (Balster, 153) BIBFRAME -same as schema.org- has the power to be machine readable, in other words, they are expressed as machine-targeted languages. As stated by DeWeese (26) standardization makes resources accessible, reusable, and repurposable. Standardization is the key for making the “Web of Data” a useful resource.

To answer the question if schema.org can accomplish what BIBFRAME intends to do there are different views. On one hand, schema.org has been expanded in a format called Schema Bib Extended taking into considerations what BIBFRAME intrinsically has. It has allowed new properties to be declared that weren't in the original schema.org. Nevertheless, some authors like Fons (25) affirm that “it will always be necessary to have a vocabulary that is used within libraries to exchange data at a level of detail that isn't useful on the web”,

Additionally, there seems to be an intrinsic conflict in trying to use schema.org on this two worlds of information management. The role of a search engine is to provide the information a user needs, nothing more, nothing less. It is very specific. As a matter of fact, its success depends on it. Schema.org website clearly states that “Schema.org focuses on defining the item types and properties that are most valuable to search engines. This means search engines will get the structured information they need most to improve search.” What libraries, on the

other hand, intend to achieve is to allow the maximum visibility but with a different perspective. “BIBFRAME engages in the development, management, and leverage of relationships existing resource to resource, resource to authority, and so on, and the use of controlled vocabularies” (Fons 34) Libraries do not pretend to be search engines discovering content, but providers of the latter. Libraries strength relies on “identifiers referencing and contextualizing unique entities” (Fons, 34). Librarians’ commitment is tied to the integrity of the information. For search engines the power is in specificity, for librarians is in context. It is hard to imagine that this two worlds could use the same tool, at least for now.

References

- Balster, K. (2018). BIBFRAMEing for Non-BIBFRAMERs: An Introduction to Current and Future Cataloging Practices. *Serials Librarian*, 74(1–4), 151–155.
<https://doi-org.ezproxy.fau.edu/10.1080/0361526X.2018.1428474>
- DeWeese, K. P., & Segal, D. (2015). *Libraries and the semantic web*. [San Rafael, California] : Morgan & Claypool Publishers, 2015. Retrieved from
<http://ezproxy.fau.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,url,uid&db=cat06361a&AN=fau.PDA003959893&site=eds-live&scope=site>
- Fons, T. (2016). Improving web visibility: into the hands of readers. *Library Technology Reports*, (5), 1. Retrieved from
<http://ezproxy.fau.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,url,uid&db=edsgao&AN=edsgcl.460283962&site=eds-live&scope=site>
- McCallum, S. (2017). BIBFRAME Development. *JLIS.It, Vol 8, Iss 3, Pp 71-85 (2017)*, (3), 71.
<https://doi-org.ezproxy.fau.edu/10.4403/jlis.it-12415>
- Magda El-Sherbini. (2018). RDA implementation and the emergence of BIBFRAME. *JLIS.It, Vol 9, Iss 1 (2018)*, (1). <https://doi-org.ezproxy.fau.edu/10.4403/jlis.it-12443>
- Seeman, D., & Goddard, L. (2015). Preparing the Way: Creating Future Compatible Cataloging Data in a Transitional Environment. *Cataloging & Classification Quarterly*,

53(3/4), 331–340.

<https://doi-org.ezproxy.fau.edu/10.1080/01639374.2014.946573>

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (n.d.). Representing and integrating bibliographic information into the Semantic Web: A comparison of four conceptual models. *Journal of Information Science*, 43(4), 525–553.

<https://doi-org.ezproxy.fau.edu/10.1177/0165551516650410>

About Schema.org. Retrieved from <https://schema.org/docs/faq.html#0>