**Submitted on:** 12/08/2014

# Authorities, Entities & Communities

**Theodore Fons**
Data Services & WorldCat Quality, OCLC, Dublin, Ohio, USA.
fonst@oclc.org



## Abstract:

*Linked data principles benefit from graph theory concepts. Graph theory itself models the relationships between objects. Linked data practice emphasizes the value of assertions (links) that describe an entity of value.*

*Traditional authority files managed by local or global communities of practice provide excellent input into the creation of entities described by the metadata representing library collections. Linked data theorists advocate the importance of making assertions about real world objects, whereas library authority files document the authorised form of a name or title for entry into a filing system or machine-readable index. An entity that provides the full context for a person is required for an entity of value on the web.*

*The move to entities as the objects of curation and quality management in library metadata management systems promises benefits to library workflows such as discovery, cataloguing, and web exposure. Those benefits include intuitive discovery, more efficient cataloguing, and effective exposure of library collections on the web.*

*To achieve these benefits, libraries and supporting library organisations should continue to invest in local and international authority files, continue to invest in aggregations of authority data such as VIAF and ISNI, use Schema.org markup for more effective web exposure, and experiment with the transformation of library metadata to entities instead of traditional record formats.*

**Keywords:** Authorities, Knowledge Cards, Metadata, Discovery, Cataloguing

## Introduction

The study of graph theory, or the mathematical structures used to model pairwise relations between objects, has a long tradition in mathematics. Many fields of study, including computer science have found fruitful applications of graph theory. The related area of study and practice that promises benefits to the management of metadata describing library collections is linked data. Linked data principles take advantage of graph theory concepts.

Linked data practice emphasizes the value of a series of assertions (links) that describe an entity of value.

The move to managing data at the entity level contrasts to the method of data management that emphasizes the quality of a metadata "record" following an encoding regime (cataloguing rules) and a document format such as MARC21 or UNIMARC. This move promises benefits to library workflows if the data is well-implemented and current experiments with library data "entification" are successful. Success can include more effective exposure of library data in web search engines, a more compelling and effective user experience in library discovery systems and more efficient metadata management workflows.

**Entities & Traditional Authority Files**

Traditional library authority files are a source of high quality data for Work and Person entities. The various communities of practice in France, the Netherlands, the German-speaking countries, and the Anglo-American countries of the US, UK, Canada, Australia, New Zealand, and South Africa have all developed well-organised networks that produce high quality metadata about persons, topics, and uniform forms of titles of bibliographic works. The most significant focus of these communities has been on providing authorised forms of author names for the authors of commercially published monographs. All of these communities of practice have developed well-documented practices and some have developed shared systems for the management and publication of these authority files. The benefit of this work has been uniform entries in library catalogue indexes for the same identity following the orthographic conventions of community languages and customary transliteration schemes. A related benefit is the use of syndetic structures to aid in name disambiguation and the promotion of identifiers to aid in data sharing and matching.

All of this output provides excellent input to the creation of entities representing library collections. One notable contrast to linked data theory and traditional library authority practice is that linked data theorists advocate the importance of making assertions about real world objects (a person, a work, a place, an organisation, etc.); that is providing links to authoritative statements about those things, whereas library authority files document the authorised form of a name or title for entry into a filing system or machine-readable index. The authorised form of names are useful assertions about a person, but an entity that provides the full context for a person is required for an entity of value on the web.

**Benefits for Library Workflows**

The move to entities as the objects of curation and quality management in library metadata management systems promises benefits to a number of library workflows:

**Discovery** As the search engines and ecommerce websites that feature bibliographic materials take advantage of entity-based search and display of data, users of library catalogues will bring the expectation of entity-based discovery to the library discovery systems. The search engines have introduced features like entity-based knowledge cards into their search results to provide contextual summaries of the entities that match a search. Entity-based library management systems provide the potential to transform library discovery from a systems that provide record-based search and display to entity-based search and display, and the introduction of knowledge cards to library discovery. Such systems can also provide a more intuitive user experience and the potential for serendipitous discovery as users follow the links and are directed to related nodes on the graph.

**Cataloguing** Entity-based cataloguing gives the opportunity for cataloguers to make improvements to the quality of entities and therefore the quality of end user discovery. That includes improvements to the usefulness and reliability of knowledge cards. It is also possible that improvements to one level of the entity graph can be shared with other levels and made manifest in the legacy record formats for legacy library systems. For example, an improvement to the work entity could provide quality improvements to many manifestation entities; and that can be reflected in traditional records created for legacy library management systems.

**Web Exposure** If libraries focus on the principle of first modelling things of interest to the web (Works, Persons, Manifestations), and expose the data in ways that are familiar to the web (such as Schema.org mark-up), then there is the potential for improvements to exposure of commodity and unique collections on the web search engines.

**The Future**

To achieve the benefits of an entity-based data model, libraries should pursue a program of transformation from the record based model. This will require investment, experimentation and collaboration. Some key elements of that program are:

- Continued **investment in local and international authority files** and cooperative management of authority files. The success of the cooperatively managed authority files by national libraries and communities with shared languages is clear and libraries should sustain the investment in these cooperative activities.

- Continued **investment in aggregations of authority data** such as VIAF and ISNI to add value to authority data. These aggregations also make authority data available to the web, improve the cataloging workflow, and disseminate canonical identifiers.

- Promote the use of expressions of the data that are already used on the web. The use of **Schema.org markup** is the method preferred by the commercial search engines and libraries should promote extensions to Schema.org to provide the detail and nuance required for the expression of library data on the web.

- **Experiment with entities.** Data quality will improve as libraries experiment with the entity model. We will find new methods for effectively exposing library collection metadata and we will require innovations in library workflow applications such as cataloguing tools and data mining techniques as we experiment with entity data.

This kind of evolutionary change, allowing for learning and refinement along the way, promises a path toward a strong future for libraries on the web.

**Acknowledgments**

I am grateful to Glenn Patton and Boaz Nadav-Manes for comments and recommendations to improve the quality of this paper.

**References**

Google: Official Blog. "Introducing the Knowledge Graph: Things, Not Strings" http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html

Hickey, Thom B. and Toves, Jenny A. “Managing Ambiguity in VIAF.” *D-Lib Magazine* 20 (2014). http://www.dlib.org/dlib/july14/hickey/07hickey.html

Svensson, Lars G. “Are Current Bibliographic Models Suitable for Integration with the Web.” *Information Standards Quarterly* 25 (2013): 7-12.

Wikimedia. (n.d.) “Graph Theory.” http://en.wikipedia.org/wiki/Graph_theory