INFORMATION STANDARDS QUARTERLY
WINTER 2013 | VOL 25 | ISSUE 4 | ISSN 1041-0031

ISQ

TOPIC

EVOLUTION OF BIBLIOGRAPHIC DATA EXCHANGE

ARE CURRENT BIBLIOGRAPHIC MODELS SUITABLE FOR INTEGRATION WITH THE WEB?

A TRANSFORMATIVE OPPORTUNITY: BIBFRAME AT GEORGE WASHINGTON UNIVERSITY, AN EARLY EXPERIMENTER

INTERVIEW WITH GILDAS ILLIEN , BnF

SCHEMA BIB EXTEND

NISO
How the information world CONNECTS

Richard Wallis

Schema Bib Extend

RICHARD WALLIS

In the evolving world of the web, bibliographic resources have gained a reputation for being difficult to discover. Search engines are on a mission to identify *things* on the web, as against just indexing texts about those things. Their initiatives could help solve some of the lack of visibility and discoverability issues in the bibliographic domain—a domain where describing things in text, as opposed to data, is the centuries old, modus operandi.

To take best advantage of such progress, you need to be part of, or at least be well represented in, the evolution of the standards and practices that are building the *things* based view of the world. This is where Schema Bib Extend fits, an influencer recognizing the concerns, experience, knowledge, and ambitions of the bibliographic corner of the web. A corner with much to offer that could be undervalued if we do not speak up and get involved.

What is Schema Bib Extend?

Schema Bib Extend is a W3C Community Group focused on establishing a consensus within the bibliographic community around proposals to submit to the WebSchemas Group for extending the Schema.org vocabulary to enhance its capabilities in describing bibliographic resources.

That statement needs unpacking:

A W3C Community Group is an open forum, without fees, where web developers and other stakeholders develop specifications, hold discussions, develop test suites, and connect with W3C's international community of web experts. The Schema Bib Extend group was formed as a Community Group to take advantage of the open forum for stakeholders the W3C provides.

The Schema.org vocabulary was developed through cooperation between Google, Bing, Yahoo! and Yandex, and released in June 2011. The purpose is to provide a vocabulary for marking up structured data on the web that will be recognized by the major search engines. The process for commenting upon and

proposing extensions and enhancements to the Schema.org vocabulary is also handled through a W3C Community Group—*WebSchemas*—with its associated *public-vocabs* mailing list.

In October 2012, I established and became chair of the Schema Bib Extend Community Group (SchemaBibEx). It has attained a membership in excess of 80 people, acting as individuals and/or representing organizations with interests in the bibliographic domain. Organizations represented include several national libraries, library system vendors, publishers, W3C, universities, cooperatives, and consortia. The group meets regularly by conference call and, via the community wiki, has already formed and submitted several proposals on topics such as Collections, Citations, and AudioBooks to the *WebSchemas* Group.

Formation of SchemaBibEx

I formed the group following many conversations that were stimulated by the release of open linked data in OCLC's WorldCat, using Schema.org as the vocabulary for data description.

By adding Schema.org-described metadata to the WorldCat pages, using the RDFa formatting technique, OCLC made available linked data descriptions of the over 300 million resources referenced in WorldCat. Schema.org was chosen as the vocabulary because of its general acceptance across the web and the fact that major search engines would recognize it. In the process of preparing these descriptions, it became clear that Schema.org did not cover certain concepts and format types. The OCLC developers created a prototype library vocabulary to supplement Schema.

In discussions, it was clear to me that there was a potential consensus that Schema.org could form the basis for describing bibliographic resources on the web, but it would need some enhancement to realize that possibility.

Following the lead of those behind Schema.org, the open group was formed, with the help of the W3C, believing that a proposal from a group of interested parties could carry more weight than those from individuals alone. Also such a group could bring informed discussion and use cases to bear on the proposals in their formation.

A change in thinking

In the early months of the group's discussions, it became clear that proposing extensions to an established general-purpose vocabulary is very different than creating and maintaining a vocabulary/standard focused on a single domain such as libraries.

Our experience and practice over many years has conditioned us to be a bit too deep and too bibliographic specific. The initial effect of this was to suggest that there was to be a significant amount of effort to identify many bibliographic vocabulary terms not present in Schema.org.

A change in approach evolved. Issues were addressed and explored by taking the Schema.org vocabulary as is and using it to describe resources, and their relationships, in the bibliographic domain. In this process, example webpages for bibliographic resources were examined to see what Schema.org markup would be appropriate. The outcome of this approach was to realize how good Schema.org was already for describing our resources, and to identify specific gaps in coverage—it had no Audiobook class for instance.

In a few cases, where the initial presumption was that new classes/properties would be required, it became clear that advice, documentation, or examples would be sufficient. In other cases, proposed tweaks to the descriptions in Schema.org documentation would be all that is needed.

CONTINUED »

In this process, example webpages for bibliographic resources were examined to see what Schema.org markup would be appropriate. The outcome of this approach was to realize how good Schema.org was already for describing our resources, and to identify specific gaps in coverage—it had no Audiobook class for instance.

An approach for Holdings

A good example of all the above is the work the group is currently engaged with to describe library holdings. This would enable libraries to describe, using Schema.org, the availability of items to loan or access in other ways.

Initial thoughts could have resulted in proposals for library-specific classes and properties. However the use of the Schema.org *Offer* class—with some adjustments to its documentation descriptions to take into account that offers can be to loan and share, as well as to sell—will go a long way to satisfying the library, available to access, use case. What then remained was some finer detailed work on which, and if any, new properties could be used to describe library-specific things such as shelf marks, call numbers, etc.

A Group with a short future

When setting up the Group, I expressed the ambition for it to have a lifetime measured in months not years. The reasoning behind this being that it was being set up to guide and inform the wider web community, served by Schema.org, on how to improve its representation of bibliographic resources—not to become a group emulating and duplicating metadata standards.

Although there is much to do, it could be possible for the majority of the issues to be addressed before the completion of the Group's second year.

What will the SchemaBibEx legacy be?

As a group representing many voices in the bibliographic domain, it has already become one looked to and referenced in broader discussions on the main, *public-vocabs*, Schema.org mailing list. Several group members are active on that list as individuals, participating in discussions some of which overlap with those in the SchemaBibEx Group.

Obviously if the Group achieves its goal, Schema.org will be better suited for the general representation of bibliographic resources, and hence such resources should be better represented in the web of data and easier to discover.

The documentation and examples that the group produces as part of its discussions could provide guidance to those wishing to describe bibliographic resources on ways to approach the issue. This should help deliver some consistency of output across the domain.

It is also apparent that through the activities of the group, system developers have been encouraged to look to using this approach to describing library resources on the web. For example, in addition to OCLC's WorldCat developments, open source library systems such as Evergreen and Koha have implemented code to expose Schema.org in their user interfaces.

If the Group achieves its goal, Schema.org will be better suited for the general representation of bibliographic resources, and hence such resources should be better represented in the web of data and easier to discover.

In summary, the SchemBibEx Group and its proposals as adopted should result in bibliographic resources being more consistently and more often represented in the web of data, and hence more discoverable. | IP | doi: 10.3789/isqv25no4.2013.06

RICHARD WALLIS (richard.wallis@oclc.org) is Technology Evangelist with OCLC in the UK.

RELEVANT LINKS

Evergreen ILS
evergreen-ils.org

Koha ILS
www.koha.org

public-vocabs W3C mailing list
lists.w3.org/Archives/Public/public-vocabs/

RDFa
www.w3.org/standards/techs/rdfa#w3c_all

Schema Bib Extend
www.w3.org/community/schemabibex/

Schema.org
schema.org

WebSchemas wiki
www.w3.org/wiki/WebSchemas

WorldCat
www.worldcat.org