Extending Schema.org
A view from the SchemaBibEx Group

Richard Wallis
Technology Evangelist
OCLC

@rjw
richard.wallis@oclc.org
Who can extend Schema.org?
Who can extend Schema.org?

- Anyone can propose extension/change
Who can extend Schema.org?

- Anyone can propose extension/change
- Just propose to the public-vocabs list
  - http://lists.w3.org/Archives/Public/public-vocabs/
- Web Schemas: http://www.w3.org/wiki/WebSchemas
Who can extend Schema.org?

- Anyone can propose extension/change
- Just propose to the public-vocab list
  - http://lists.w3.org/Archives/Public/public-vocab/
- Web Schemas: http://www.w3.org/wiki/WebSchemas
- Domain interested groups carry weight
Who/What has extend Schema.org?
Who/What has extend Schema.org?

- IPTC - News
- Genealogy - historical-data.org
- E-commerce - GoodRelations
- Learning/Education - LRMI
- Medical/Health
- TV/Radio
The mission of this group is to discuss and prepare proposal(s) for extending Schema.org schemas for the improved representation of bibliographic information markup and sharing. The group will seek consensus around, and support for, proposal(s) to the W3C WebSchemas Group. This Community Group will not, itself, produce technical specifications.

Get involved!
Anyone may join this Community Group. All participants in this group have signed the W3C Community Contributor License Agreement (CLA).

No reports yet published. The Chair is responsible for publishing reports.
More about publishing…

Richard Wallis
Chair

Participants
The mission of this group is to discuss and prepare proposal(s) for extending Schema.org schemas for the improved representation of bibliographic information markup and sharing. The group will seek consensus around, and support for, proposal(s) to the W3C WebSchemas Group. The Community Group will not, itself, produce technical specifications.
Different to creating your own vocabulary
Different to creating your own vocabulary

- Schema.org is already established
Different to creating your own vocabulary

- Schema.org is already established
- It is for all domains
Different to creating your own vocabulary

- Schema.org is already established
- It is for all domains
- You may not agree with class and property names
Different to creating your own vocabulary

- Schema.org is already established
- It is for all domains
- You may not agree with class and property names
- You need to build on established classes
Different to creating your own vocabulary

- Schema.org is already established
- It is for all domains
- You may not agree with class and property names
- You need to build on established classes
- Users may not be familiar with your domain
The approach
Don’t:
Don’t:

Hunt for your favorite classes and properties missing from Schema.
The approach

Don’t:

Hunt for your favorite classes and properties missing from Schema.

Do:

Try to describe your resources using Schema to identify what is difficult/impossible
The approach

Don’t:

Hunt for your favorite classes and properties missing from Schema.

Do:

Try to describe your resources using Schema to identify what is difficult/impossible

Reality check:

Will consumers of Schema care
Linked Data
Linked Data

• 300+ Million resources
• Schema.org
• Embedded RDFa
• Links to: Dewey, LCSH LCNAF, DOI, VIAF, FAST, ...
• License: ODC-BY
• Experimental - June 2012
• Continuing development:
  - Vocabulary, Content-negotiation, Links, Works, ...
World War I

Author: H P Willmott

Summary: A visual guide to the causes, battles, and forces of World War I.

Rating: ★★★★★ (not yet rated)
RDFa and schema.org all the library things

TLDR: The Evergreen and Koha integrated library systems now express their record details in the schema.org vocabulary out of the box using RDFa. Individual holdings are expressed as Offer instances per the W3C Schema Bib Extension community group proposal to parallel commercial sales offers. And I have published a branch to give the same capabilities to the VuFind discovery layer, as well.

In the spring of 2012, I took my first steps in the structured data world by teaching Evergreen 2.2 how to express some record details in the schema.org. It was a small step towards taking the machine-readable data that we had made useful to humans on the record detail catalogue page and marking it up so that it was once again machine readable. At that time, Evergreen only knew how to map MARC data to two schema.org types (Book and MusicRecording—which should have been MusicAlbum, but I eventually fixed that) and a handful of attributes: name, ISBN, publisher, publication date, author, contributor, and keywords. Pretty barebones, but a start nonetheless.

I used the HTML5 microdata approach because I was new to structured data and microdata was what was demonstrated in all of the schema.org examples, so it seemed like the obvious choice. Over the last year, however, I realized that RDFa is a W3C standard for accomplishing the same goals as microdata, bolstered by an open community standards-making process, and featuring the ability to mix in properties and types from multiple vocabularies. I touched on this in my Evergreen 2013 conference presentation: Structured data: making metadata matter for machines. While RDFa Lite is extremely easy to get started with, I have been diving deeper into RDFa proper to make use of some of the more advanced properties, such as #about to work around unwanted chaining introduced by #href attributes.

Over the last few weeks, I was able to concentrate on improving the schema.org mapping for Evergreen--introducing holdings as instances of the http://schema.org/Offer class, providing much more granular author and contributor data--and cut over to RDFa. While the tools at RDFa Tools were quite useful for debugging my efforts, I also have to thank the denizens of the #rdfa IRC channel (Manu Sporny in particular) for patiently helping me understand some of my rookie mistakes. Ben Shum also kept me honest by patiently testing multiple iterations of my branches with the Google Rich Snippets tool and reporting any issues that he encountered; this led to my realization that using #resource and #about were necessary in some contexts.

Once I had worked out a decent mapping in Evergreen (a library system I have been contributing to for over six years now), I decided to tackle the VuFind discovery layer. VuFind uses a straightforward template system, and I was able to put together a branch that integrated schema.org as RDFa (details at VuFind bug 425), building on
TLDR: The Evergreen and Koha integrated library systems now expose vocabulary out of the box using RDFa. Individual holdings are exposed Bib Extension community group proposal to parallel commercial services and give the same capabilities to the VuFind discovery layer, as well.

I used the HTML5 microdata approach because I was new to structured data and microdata was what was demonstrated in all of the schema.org examples, so it seemed like the obvious choice. Over the last year, however, I realized that RDFa is a W3C standard for accomplishing the same goals as microdata, bolstered by an open community standards-making process, and featuring the ability to mix in properties and types from multiple vocabularies. I touched on this in my Evergreen 2013 conference presentation: Structured data: making metadata matter for machines. While RDFa Lite is extremely easy to get started with, I have been diving deeper into RDFa proper to make use of some of the more advanced properties, such as about to work around unwanted chaining introduced by @prefix attributes.

Over the last few weeks, I was able to concentrate on improving the schema.org mapping for Evergreen—introducing holdings as instances of the http://schema.org/Offer class, providing much more granular author and contributor data—and cut over to RDFa. While the tools at RDFa Tools were quite useful for debugging my efforts, I also have to thank the denizens of the #rdfa IRC channel (Manu Sporny in particular) for patiently helping me understand some of my rookie mistakes. Ben Shum also kept me honest by patiently testing multiple iterations of my branches with the Google Rich Snippets tool and reporting any issues that he encountered; this led to my realization that using @type and @about were necessary in some contexts.

Once I had worked out a decent mapping in Evergreen (a library system I have been contributing to for over six years now), I decided to tackle the VuFind discovery layer. VuFind uses a straightforward template system, and I was able to put together a branch that integrated schema.org as RDFa (details at VuFind bug 425), building on

Schema Bib Extend Community Group

The mission of this group is to discuss and prepare proposal(s) for extending Schema.org schemas for the improved representation of bibliographic information markup and sharing. The group will seek consensus around, and support for, proposal(s) to the W3C WebSchemas Group. This Community Group will not, itself, produce technical specifications.

Reports

No reports yet published. The Chair is responsible for publishing reports. More about publishing...

News

Add new post →

October Meeting

Get involved!

Anyone may join this Community Group. All participants in this group have signed the W3C Community Contributor License Agreement (CLA).

JOIN THIS GROUP

or learn how to join or request an account.

Participants

Richard Wallis
Chair
Model to help discussion
Proposals
Proposals

• Citation - Accepted
Proposals

• Citation - Accepted
• Collection/Parts - Proposed
Proposals

- Citation - Accepted
- Collection/Parts - Proposed
- AudioBook - New type
Proposals

• Citation - Accepted
• Collection/Parts - Proposed
• AudioBook - New type
• Holdings - Based on Offer
Proposals

• Citation - Accepted
• Collection/Parts - Proposed
• AudioBook - New type
• Holdings - Based on Offer
• Modify Offer (descriptions) for non-commercial view
Proposals

- Citation - Accepted
- Collection/Parts - Proposed
- AudioBook - New type
- Holdings - Based on Offer
- Modify Offer (descriptions) for non-commercial view
- Work Relationships
The mission of this group is to discuss and prepare proposal(s) for extending Schema.org schemas for the improved representation of bibliographic information markup and sharing. The group will seek consensus around, and support for, proposal(s) to the W3C WebSchemas Group. This Community Group will not, itself, produce technical specifications.

Reports

No reports yet published. The Chair is responsible for publishing reports.
More about publishing...

News

Add new post →

October Meeting

Participants

Richard Wallis
Chair