

Schema.org, FOAF and Linked Data

Lessons for Web-scale vocabulary deployment

Dan Brickley <danbri@google.com>

1. Overview

- What can we learn from schema.org, FOAF and Linked Data about variety in 'big data'?
 - Brief overview of Schema.org
 - Varieties of variety:
 - extensibility points for cross-domain vocabulary
 - identifiers, sameas and reference-by-description
 - documenting idioms and patterns
 - Current challenges: roles, actions, and enumerations.

2. Describing a Movie with schema.org

```
<div itemscope itemtype="http://schema.org/Movie">
```

```
<h1 itemprop="name">
```

Pirates of the Carribean: On Stranger Tides

```
</h1>
```

```
<span itemprop="description">
```

Jack Sparrow and Barbossa embark on a quest to find the elusive fountain of youth, only to discover that Blackbeard and his daughter are after it too.

```
</span>
```

```
</div>
```



What is Schema.org?

This site provides a collection of schemas, i.e., html tags, that webmasters can use to markup their pages in ways recognized by major search providers. Search engines including Bing, Google, Yahoo! and Yandex rely on this markup to improve the display of search results, making it easier for people to find the right web pages.

Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data. On-page markup enables search engines to understand the information on web pages and provide richer search results in order to make it easier for users to find relevant information on the web. Markup can also enable new tools and applications that make use of the structure.

A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.

We invite you to [get started!](#)

View our blog at blog.schema.org.

Schema.org partners

Google™

 Microsoft

YAHOO!

Yandex



5. Schema.org - the wider community

So far, we've seen many tremendous enhancements to schema.org come from:

- IPTC rNews group for News publishing
- GoodRelations for products and e-commerce
- W3C Government Linked Data Working Group for dataset description
- LRMI for educational resources, co-led by the Association of Educational Publishers and the Creative Commons
- Many others, both individuals and organizations

see schema.org site for links to the W3C Web Schemas group, vocab proposals and more



6. Schema.org markup in practice

- **Thing**

- CreativeWork
 - Article
 - ScholarlyArticle
 - MedicalScholarlyArticle
 - Blog
 - Book
 - ExercisePlan
- Event
 - ChildrensEvent
 - Festival
 - UserInteraction
 - UserBlocks
 - UserComments
 - UsersDownloads
- Organization

- **Thing**

- Person
- Place
 - CivicStructure
 - Airport
 - EventVenue
 - GovernmentBuilding
 - CittyHall
 - Embassy
 - LocalBusiness
 - FinanicalService
 - BankOrCreditUnion
- Product
 - IndividualProduct
 - ProductModel
 - SomeProducts
- Property



12. Vocab pipeline (W3C WebSchemas...)

Proposals and Discussion

[\[edit\]](#)

These proposals are working towards inclusion in [changes.html](#), the file that serves as a log of changes, and for communicating requested changes to the schema.org administrators.

Topic	Status	Summary	Vocab	Comments
Citation Promotion	Proposal	To promote the 'citation' property currently on MedicalScholarlyArticle up to CreativeWork .	Moves citation up to CreativeWork	
Accessibility	Proposal	Extension for describing accessibility properties of a CreativeWork .		Draft published to elicit feedback.
Add Organization department property	Accepted	Add one property 'department' to Organization, to allow local businesses to describe their depts (e.g. opening hours) in detail. Final discussion on whether to call it subOrganizationOf instead.	department property (range and domain both Organization)	Proposal very greatly simplified after WebSchemas discussion.
Add Quotation type	Proposal	Add one type, 'QuotationSchema', a kind of CreativeWork, alongside some supporting properties.	Quotation, spokenByCharacter, incorrectlyAttributedTo, aboutEvent, fromEvent, address.	Several other requirements are addressed by re-use of existing properties. Debate about whether incorrectlyAttributedTo is worthwhile. Some details to finalise.
sameAs	Proposal	Property to indicate that a single identity is being described.	Essentially same meaning as owl:sameAs, but suggesting a distinct name to make it easier to talk about; also schema.org data is typically scruffier than OWL re http-range-14.	Draft published to elicit feedback.
Actions in Schema.org	Proposal	A proposal for a new Action type.	See full entry and proposal for details. May2013 version builds on two previously published designs.	The schema.org team solicit comments on this draft via public-vocabs@w3.org.
Orders in Schema.org	Proposal	"While schema.org already supports marking up offers to sell products (pre-transaction), it does not currently provide schema for the common kinds of confirmations and notifications around orders (post-transaction)".	Adds classes OrderStatus, Order, DeliveryMethod plus associated properties.	Complements Actions proposal, but independent.
Reservations in Schema.org	Proposal	This proposal includes new types for describing reservations and tickets	Adds classes Reservation, ReservationStatus, ProgramMembership, Flight, TrainTrip, TrainReservation, FoodEstablishmentReservation, FlightReservation, EventReservation, ReservationPackage, BusReservation, BusTrip, RentalCarReservation, RentalCar, Car, LodgingReservation plus associated properties.	Complements Actions proposal, but independent.
EmailMessage type	Proposal	This proposal adds a single new type, EmailMessage	EmailMessage is a CreativeWork representing email messages, just as WebPage already represents Web pages.	No properties currently defined; proposals welcomed.
Workers Union type	Accepted	This proposal adds a single new type, WorkersUnion, addressing issue 17 in the WebSchemas tracker.	"A Workers Union (...) is an organization that promotes the interests of its worker members by collectively bargaining with management, organizing, and political lobbying."	One type, see draft schema .
TV and Radio proposal	Accepted; published in next version	Proposes modest changes and additions to support TV and radio (from EBU and BBC).	Adds Series, Season, Episode under CreativeWork. Existing TVSeries, TVSeason, TVEpisode under Series/Season/Episode. Adds RadioSeries/Radio/Season/Radio/Episode alongside. Adds Programme/Clip for programmes other than episodes. Adds description of a service with Service/PublicationEvent for broadcast, streaming, VoD, etc.	See also background notes and [11] . Draft initially from EBU/BBC/NoTube discussions. Nearby vocab: see VideoObject , AudioObject , Movie , Event re scheduling. See also UserPlays , a kind of UserInteraction . Comics and TV/Radio also share a concern for describing fictional characters and narratives .
Sports	Some initial discussions.	Several parties are working towards a Sports Vocabulary proposal. See query from Paul Kelly (SportsML) , Microdata + Sports Stats post describing earlier ESPN/Google collaboration, and the Sports Ontology used by BBC, Press Association and others. See also sports-schema-collab mailing list.	Several sources including a full draft from the ESPN+Google work. Needs summary here.	March 2nd - posted writeup of the ESPN/Google Microdata vocabulary; next is discussion on finding common basic sports vocab for schema.org.
Comics and Serials	Discussing Proposals	Proposal from Marvel (see examples).	Adds PeriodicalSeries (under Intangible(?)), PeriodicalIssue (under CreativeWork), Comic Issue (under PeriodicalIssue), GraphicNovel (under Book). Approx 6-9 properties for each of these classes.	Generally well received and specified, but some detailed discussion re fine-grain detail of comics (imprint, page count etc.). Can those properties be shared with Book/numberOfPages , or ScholarlyArticle ? Should PeriodicalSeries be intangible?
Event schema update	Hopefully Final Proposal - nreponse.	Proposes some modifications to Event based on deployment experience.	Adds eventStatus, previousStartDate, previousEndDate; eventCategory; makes startDate/endDate repeatable. Encourages use of 'url' from Thing .	Fairly modest proposals based on implementor feedback. May2013 update addresses the last open issue on repeating events.

13. Comparisons

Schema.org and previous RDF approaches.

- "classic" RDF imposes vocabulary-variety on publishers: 100s of vocabularies to choose from.
- classic Linked Data expects publishers to reconcile and link diverse identifiers.
- Schema.org: greater diversity of publishers - now over 5 million domains.

14. Extensibility points for x-domain data

Schema.org provides a shallow vocabulary covering many topics. How to go deeper?

- Syntax is essentially RDF (microdata/json-ld/rdfa); additional detail can be provided using any RDFS vocabulary.
- A role for 'external enumerations', e.g. in PlaceOfWorship we can't sub-type for every religion. Exploring bridges to Wikipedia/wikidata.

15. Identifiers

- Early FOAF "Linked Data" smushed together entity descriptions via common identifying properties.
- Since 2006, Linked Data movement advocated for "URIs for everything"
- Synthesis: Schema.org is reference-by-description plus <<http://schema.org/sameAs>> links to well known indicative pages.

16. Idioms and Patterns

On the scale of the Web, there will always be different levels of expressive detail, and potential for several ways to say same thing.

- Validators need to document and match specific patterns needed in applications.
- Inevitable to have different ways of saying same thing, e.g. `openingHours`, `OpeningHoursSpecification`.

2. Current Challenges

- Roles: yesterday we announced a draft design to squeeze more expressivity into the simple flat RDF model, e.g. temporal qualifiers.
- Actions: we have "nouns", working now on a model for description "action opportunities", i.e. verbs.
- Datasets: while simple binary relations (RDF structured data) is useful, it can be oversold. Added Dataset type, engaging with W3C via CSV on the Web WG to bring tabular/CSV data into an integrated picture. e.g. parking spaces in a CSV file.

17. Questions

...

(or to danbri@google.com)