

W3C Open Data Workshop

The schema.org experience

Dan Brickley <danbri@google.com>

Introducing schema.org

- Search engines collaborate to improve the Web
 - Bing, Yahoo!, Yandex and Google working together
 - Launched in June 2011
 - Large vocabulary for use in Web page markup
 - hundreds of types of thing & their properties
 - evolving via W3C-hosted open community
 - broad scope, practical focus
 - Based on W3C standards (RDFa, HTML5Microdata)
- Builds on ideas from Semantic Web, Microformats and Linked Data community

```
<div itemscope itemtype="http://schema.org/MedicalScholarlyArticle">
  <link itemprop="audience" href="http://schema.org/Clinician"/>
  <meta itemprop="publicationType" content="Meta-Analysis"/>
  <h1><span itemprop="name">New guidelines for metformin and
    diabetes mellitus</span></h1>
  <p>
    <span itemprop="author" itemscope itemtype="http://schema.org/Person">
      <span itemprop="name">Dr. John Smith</span>
      <br><span itemprop="affiliation">Medical University</span>
    </span>
    <br><span itemprop="datePublished">2012-03-24</span>
  <p><b>Abstract:</b>
  We review clinical evidence related to the use of metformin for
  treatment of type-2 diabetes mellitus and provide new clinical guideline
  recommendations.</p>
  <p><b>MeSH subject headings:</b>
  <span itemprop="about" itemscope itemtype="http://schema.org/Drug">
    <span itemprop="name">Metformin</span>
    <span itemprop="code" itemscope
      itemtype="http://schema.org/MedicalCode"
      itemid="http://www.ncbi.nlm.nih.gov/mesh/D02.078.370.141.450">
      <!-- Note: use of itemid is not mandatory, but recommended when an
        external enumeration is available -->
      <meta itemprop="code" content="D02.078.370.141.450"/>
      <meta itemprop="codingSystem" content="MeSH"/>
    </span>
  </span>;
  <span itemprop="about" itemscope itemtype="http://schema.org/MedicalCondition">
    <span itemprop="name">Diabetes Mellitus, Type 2</span></p>
    <span itemprop="code" itemscope
      itemtype="http://schema.org/MedicalCode"
      itemid="http://www.ncbi.nlm.nih.gov/mesh/C18.452.394.750.149">
      <!-- Note: use of itemid is not mandatory, but recommended when an
        external enumeration is available -->
      <meta itemprop="code" content="C18.452.394.750.149"/>
      <meta itemprop="codingSystem" content="MeSH"/>
    </span>
  </span>
  ...
  <h3>Guidelines</h3>
  <span itemscope itemtype="http://schema.org/MedicalGuidelineRecommendation">
    Recommendation: we recommend monotherapy with
    <span itemprop="subject">metformin</span> as an initial pharmacologic therapy
    to treat most patients with <span itemprop="subject">type 2 diabetes</span>
    (Grade: <span itemprop="recommendationStrength">strong recommendation</span>;
    <link itemprop="evidenceLevel" href="http://schema.org/EvidenceLevelA">high-quality
    evidence</link>).
  </span>
</div>
```

Example uses at Google

[Little Water Cantina - Eastlake - Seattle, WA](#)
www.yelp.com › Restaurants › Mexican
★★★☆☆ 90 reviews - Price range: \$\$
90 Reviews of **Little Water Cantina** "Three things are on my list when I eat out: great food, atmosphere, and

[Vegetarian \(Vegan Pizza No Cheese\) Recipe - Food.com - 248865](#)
www.food.com/recipe/vegetarian-vegan-pizza-no-c...
★★★★★ 2 reviews - 1 hr 32 mins - 242.9 cal
Aug 26, 2007 – This is from my dad, who developed some **vegan recipes** that don't have any cheese, and you

[Leonard Cohen – Free listening, videos, concerts, stats, & pictures at...](#)
www.last.fm/music/Leonard+Cohen
Watch videos & listen to **Leonard Cohen**: Suzanne, Hallelujah & more, plus 132 pictures. **Leonard Cohen**, (born September 21, 1934 in Montréal, Quebec, ...

Track	Duration
Suzanne	♫ 3:48
The Darkness	♫ 4:29
Going Home	♫ 3:51
Hallelujah	♫ 6:12

- Enhanced search results - 'Rich Snippets'
- Custom Search Engines (for jobs, education, agriculture, datasets...)
- Working to improve description, discovery & integration of open datasets via schema.org

Thing > CreativeWork > Article > NewsArticle

A news article

Property	Expected Type	Description
Properties from Thing		
additionalType	URL	An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax – the 'typeof' attribute – for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.
description	Text	A short description of the item.
image	URL	URL of an image of the item.
name	Text	The name of the item.
url	URL	URL of the item.
Properties from CreativeWork		
about	Thing	The subject matter of the content.
accountablePerson	Person	Specifies the Person that is legally accountable for the CreativeWork.
aggregateRating	AggregateRating	The overall rating, based on a collection of reviews or ratings, of the item.
alternativeHeadline	Text	A secondary title of the CreativeWork.
associatedMedia	MediaObject	The media objects that encode this creative work. This property is a synonym for encodings.
audience	Audience	The intended audience of the item, i.e. the group for whom the item was created.
audio	AudioObject	An embedded audio object.
author	Organization or Person	The author of this content. Please note that author is special in that HTML 5 provides a special mechanism for indicating authorship via the rel tag. That is equivalent to this and may be used interchangeably.
award	Text	An award won by this person or for this creative work.
awards	Text	Awards won by this person or for this creative work. (legacy spelling; see singular form, award)
comment	UserComments	Comments, typically from users, on this CreativeWork.
contentLocation	Place	The location of the content.
contentRating	Text	Official rating of a piece of content—for example, 'MPAA PG-13'.
contributor	Organization or Person	A secondary contributor to the CreativeWork.

learningResourceType	Text	The predominant type or kind characterizing the learning resource. For example, 'presentation', 'handout'.
mentions	Thing	Indicates that the CreativeWork contains a reference to, but is not necessarily about a concept.
offers	Offer	An offer to sell this item—for example, an offer to sell a product, the DVD of a movie, or tickets to an event.
provider	Organization or Person	Specifies the Person or Organization that distributed the CreativeWork.
publisher	Organization	The publisher of the creative work.
publishingPrinciples	URL	Link to page describing the editorial principles of the organization primarily responsible for the creation of the CreativeWork.
review	Review	A review of the item.
reviews	Review	Review of the item (legacy spelling; see singular form, review).
sourceOrganization	Organization	The Organization on whose behalf the creator was working.
text	Text	The textual content of this CreativeWork.
thumbnailUrl	URL	A thumbnail image relevant to the Thing.
timeRequired	Duration	Approximate or typical time it takes to work with or through this learning resource for the typical intended target audience, e.g. 'P30M', 'P1H25M'.
typicalAgeRange	Text	The typical range of ages the content's intendedEndUser, for example '7-9', '11-'.
version	Number	The version of the CreativeWork embodied by a specified resource.
video	VideoObject	An embedded video object.
Properties from Article		
articleBody	Text	The actual body of the article.
articleSection	Text	Articles may belong to one or more 'sections' in a magazine or newspaper, such as Sports, Lifestyle, etc.
wordCount	Integer	The number of words in the text of the Article.
Properties from NewsArticle		
dateline	Text	The location where the NewsArticle was produced.
printColumn	Text	The number of the column in which the NewsArticle appears in the print edition.
printEdition	Text	The edition of the print product in which the NewsArticle appears.
printPage	Text	If this NewsArticle appears in print, this field indicates the name of the page on which the article is found. Please note that this field is intended for the exact page name (e.g. A5, B18).
printSection	Text	If this NewsArticle appears in print, this field indicates the print section in which the article appeared.

This class contains derivatives of IPTC rNews properties. rNews is a data model of publishing metadata with serializations currently available for RDFa as well as HTML5 Microdata. More information about the IPTC and rNews can be found at rnews.org.

WebSchemas group at W3C: public-vocabs@w3.org

Proposals for Schema.org

see also [changes.html](#) for pending and recent changes at schema.org.

This page attempts to offer a canonical list of public proposals for [schema.org vocabulary](#) extensions and improvements. We also have an [issue tracker](#) to monitor concrete actions, bugfixes and workflow, and the public-vocabs@w3.org W3C mailing list for discussions. Sometimes proposals evolve from offlist and face-to-face discussions; when possible advance warning of 'expected' or possible proposals can be indicated here. ([announced 24 Feb 2012](#)).

- Editor: Dan Brickley (danbri@danbri.org).
- Status: Feel free to update the table for accuracy and detail, but discuss status-oriented changes with the editor to avoid confusion.
- If you add a proposal, please also mail the public-vocabs@w3.org list to start a discussion (and link it from here)

Proposals should include *markup examples* as well as vocabulary definitions, and should point to related and prior work. It is also important to give examples of real sites that have suitable content. See [Posting markup examples in W3C wikis](#) documentation for details on how to post examples here.

Schema.org proposals can take any form during discussions, but ultimately they should take the form of HTML+RDFa schema definitions; some real examples are in the webschemas area of the [W3C Mercurial repository](#).

Note: When preparing a proposal, we need to know for each type and property, its name and description, alongside examples of usage. There are lots of different ways in which these can be prepared: wiki pages, text files, spreadsheets, PDFs etc. Do note that each property is considered as a global entity, ie. a property called 'foo' gets just one description - you can't give different descriptions for its use with the "Person" type, versus its use with the CreativeWork type.

To search just these proposals, the issue tracker and schema.org-related mailing lists, try a [this custom search](#).

Proposals and Discussion

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
Accessibility	Addition candidate	Extension for describing accessibility properties of a CreativeWork .	Accessibility	Draft published to elicit feedback.
Learning Resources	Addition Candidate drafted , based on LRMI ID submission - please review .	The Learning Resources Metadata Initiative has drafted a set of terms to improve schema.org's use with educational materials.	Mainly adds new properties (typically but not necc. on CreativeWork): intendedEndUserRole , educationalAlignment , with values in a new class, AlignmentObject ; educationalUse (Text values), timeRequired (a Duration); typicalAgeRange (Text valued); interactivityType (Text); learningResourceType , ...	Comments: see version 1.0 and discussion list .
Datasets	Addition Candidate drafted	A proposal extending schema.org for describing datasets and data catalogs.	Adds the Dataset , DataCatalog , and DataDownload types and supporting properties.	An example page with dataset microdata is available here . For related efforts, see the W3C Government Linked Data Working Group .
Tech Article	Addition Candidate drafted	A proposal extending schema.org for describing technical content.	Adds dependencies, task, and technicalAudience. Adds descriptions for genre (CreativeWork) and about (CreativeWork).	Addition candidate
API Reference	Addition Candidate drafted	A proposal extending schema.org for describing Application Programming Interface (API) reference content.	Derived from TechArticle; adds assembly, assemblyVersion, programmingModel, targetPlatform	Addition candidate
Code	Addition Candidate drafted	A proposal extending schema.org for describing human readable source code.	Adds codeRepository , codeSample , programmingLanguage , runtime , targetProduct .	Addition candidate
Audience	Addition Candidate drafted	A proposal extending schema.org for describing audiences, building around http://schema.org/Audience type, also used by (and coordinated with) LRMI and MedicalHealth vocabulary.	Add property audience to type CreativeWork (done) and Product . Also ParentAudience , ... see draft for details.	Partially included already; final tweaks.
Activities and Actions	Proposal	A proposal for new classes Activity and Action , possibly also addressing some issues around UserInteraction .	Adds classes Activity , Action , AggregateActivity plus associated properties, and some recommendations on changes to related existing vocabulary. See April 2012 announcement ; Reworked Nov 2012 .	This addresses some known problems with the current UserInteraction design, and integrated activitystream -like functionality into schema.org .
TV and Radio proposal	Addition Candidate - please review!	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/RadioSeason/RadioEpisode 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 [1] . 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 writup 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	Proposal - please review	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. Some examples needed, especially around startDate/endDate ; it is obvious how they work with multiple values! <i>Still needs work</i> . Some pieces show up in Actions/Activities proposal now.
Historical Data proposal	Proposal	The historical-data.org site proposes some extensions to schema.org to address historical and genealogical (family history) scenarios.	Adds HistoricalRecord , HistoricalEvent (subclass of Event), overrides attendees, subEvents , superEvent to point to HistoricalEvent .	There is an associated implementation, github repository , and blog .
Biological Databases	Proposal	A schema extension for describing biological databases, Sparce .	Adds a class 'BiologicalDatabaseEntry' as a kind of CreativeWork , introducing 'entryID', 'isEntryOf', 'taxonID'. Adds 'BiologicalDatabase' also subclass of CreativeWork , with no special properties. Both also use 'breadcrumb' from WebPage .	Others have also mentioned interest in adding some notion of species.
Breadcrumbs	Proposal	A proposal resolving current problems handling breadcrumbs.	Adds Breadcrumb type with child property and supposes to use this class on breadcrumb property of a WebPage . An important schema, but needs attention. See discussion on using literal values.	
Discussion	Proposal	A proposal extending schema.org for describing online discussion forums and message boards.	Adds attachmentURL , contributorType , discussionState , interestedUsers , postCount , postType , votedHelpful .	Addition candidate
Innovation	Proposal	A proposal extending schema.org for describing innovation and related concepts. Discussion found here: [2]	Covers concepts and terms such as Need, Benefit, Usage, Disruption, Innovator, Innovation, Embodiment, to name a few. See [3] for complete list of terms.	Addition candidate
...				
<i>Proposal name</i>	<i>How far along</i>	<i>Summary of contribution.</i>	<i>Overview of new terms</i>	<i>Comments on status, particularly regarding integration, overlaps and open issues.</i>

Brainstorming, Use Cases and Advance Notice

Datasets

- Please read our position paper!
- We've added types for Dataset, DataDownload, DataCatalog
 - thanks to Jim Hendler & team, W3C DCAT, ...
- Allows any/all schema.org properties to be used for dataset discovery
 - but what about the information in the data? can we find ways to *look inside* datasets?
- Where do graphs stop and tables start?
 - fulfilling RDF's original promise as a metadata system
- Collaborations: W3C Community groups?