Schema.org / Web of Things

22 Sept 2016, W3C TPAC, Lisbon

Dan Brickley <danbri@google.com>
Schema.org CG chair / project webmaster

Schema.org and WoT/IoT

- Quick overview of Schema.org
  - Origins, approach and status
  - Approach to standardization
  - (Early) thinking on schema collaboration for IoT/WoT
  - Discussion

Schema.org overview: origins, approach and status

- Q: **What is it?**
- A: Large (and growing) collection of descriptive schemas, used primarily in the public Web, but also in email messages and other contexts.
- Q: **What kinds of descriptions?**
- A: Many! People, places, things, events, recipes, reviews, products, businesses, opening hours, creative works (books, articles etc.), many more.
- Q: **Where is it used?**
- A: Initially and primarily by search engines, but also sites like Pinterest, or iniatives like http://api.science.ai/ (scholarly publishing)
- Q: How big is it?
- A: Core vocabulary currently has 583 types, 846 properties, and 114 Enumeration values. More in extension sections.

Approach (continued…)

- Q: What **technical approach** / syntax / standards does it build on?
- A: Launched in 2011 using HTML5 Microdata, later embraced also RDFa and JSON-LD as alternate syntaxes. Underlying approach is ~ lightweight RDF.
- Q: **How widely used is it**?
- A: Very. Used in a significant proportion of the Web. Generally publication and consumption are on a page-by-page basis (rather than sites as graphs).
- Q: How broad is the **scope**?
- A: From the start, as broad as Web search - any topics that are substantially represented in the public Web, search queries and applications. In ~2013 we expanded also to explore schemas in email, covering more personal data.

Schema.org: Approach to standardization

- Founded as a collaboration between Google, Bing and Yahoo in 2011, and soon joined by Yandex.
- Moved quickly to using W3C Semantic Web Interest Group's public-vocabs@w3.org forum as its public community.
- Builds on (and helps deploy) W3C standards, but not itself a W3C standard or WG.
- In recent years, established an informal process that combines W3C Community Group discussion with final review of regular releases by a small steering group with reps from original founders and the wider community.
- Changes, additions and improvements come from public discussions and occur primarily on the Github site.

Schema.org: Standards approach, continued

- Initial approach differed from "classic RDF" (Linked Data, Semantic Web etc.) by having a much larger vocabulary. Instead of 40-50 terms per vocabulary, began at around 500 and has since doubled.
- In 2015 introduced a soft modularization mechanism, 'hosted extensions' (bib.schema.org, auto.schema.org, health-lifesci.schema.org) as well as a distributed approach to independent extension, e.g. see http://gs1.org/voc/
- Various tradeoffs. Stability and agility; usability, monolithic-icity; maintainability, modularity, etc.
- Each release is also published as a named frozen snapshot, including definitions in RDF-based formats, to support W3C spec citation.

Schema.org and IoT / WoT

- Aside: we tend to say "IoT" since "WoT" evokes "Web of Trust"; but we operate at a level of abstraction closer to W3C's WoT than to "technical plumbing".
- Current draft release (for 3.2) includes an initial IoT "hosted extension", see http://iot.webschemas.org/
  - Initially a position paper to establish perspective, and a new mailing list (please join!)
- Rough approach (including some personal perspective):
  - IoT (and WoT) has a **huge scope**.
  - In terms of descriptive schemas, the *Web/Internet* component may be less important than the ***real world Things*** angle. We have a lot of vocabulary for real world things.
  - The landscape has many many standard initiatives. We seek concrete usecases where we can make ourselves (and our schemas) useful, and plug into other efforts, including W3C WoT.
  - Position paper gives some brief scoping usecases. The existing use of schema.org within email may provide a model for thinking about security aspects (i.e. left to surrounding context).
  - **Schema.org doesn't do protocols**, except "Action" has "potentialAction". Not clear how deep to go!

Usecase areas from position paper

- User data portability (e.g. my Withings scale records my weight, can factual data records / datasets be made available to users and apps in a standard way?).
- Beacons and the description of the physical environment, including sensors and sensor-based datasets - e.g. PhysicalWeb, physical accessibility, opening hours, transport data etc. (building on existing Dataset type?)
- Smart smart Assistants - e.g. personal data (flights), potential actions, …
- On-device content. E.g. TV/video content, music, personal/family photos and media archives. Schema.org has good models for TV/Radio, MusicBrainz etc.
- Energy efficiency, ...

Discussion

See https://www.w3.org/2016/09/22-wot-minutes.html for WoT WG minutes / notes from presentation.