

Vocabulary management plans at W3C

Vocabulary preservation session, DC 2013

2013-9-3

Ivan Herman, W3C

The past...

- ▶ For a long time, W3C did not *really* deal with vocabularies
- ▶ The philosophy was: this is a distributed world, W3C should not “centralize” vocabulary management
- ▶ And that was backed up by some developments:
 - Dublin Core
 - FOAF
 - Creative Commons (cc)
 - SIOC

But... there are issues

- ▶ Development and maintenance, versioning, etc, of a vocabulary is often a matter of collaboration
 - an environment should be available for that
- ▶ Maintaining a proper web site may be a challenge on long term
- ▶ Stability of the vocabulary URI is essential
- ▶ *Preservation of the vocabulary on long term is essential*
- ▶ Finding a specific vocabulary is a major challenge

These requirements make it difficult for groups to develop new vocabularies

W3C's Current Approach

1. Provide an environment where vocabulary development and maintenance can happen
2. Offer a vocabulary hosting service
3. Provide a vocabulary directory

1. Vocabulary developments

- ▶ W3C already has the “Community Group” structure:
 - anybody can start a CG if there are 3 others who back it up
 - no strings attached, no membership required
 - W3C offers archived mailing lists, wiki, blog
 - minimal legal protection (essentially a RF and Patent Free environment)
- ▶ Community Groups are already a perfect place for the development of new vocabularies by various groups out there
- ▶ But this fact is not really well-known ☹️

Example: Open Annotation Ontology

The screenshot shows a web browser window with the URL www.w3.org/community/openannotation/. The page features the W3C logo and the text "W3C Community and Business Groups". A search bar is located in the top right corner. The main content area is titled "Open Annotation Community Group" and contains a paragraph describing the group's purpose: "The purpose of the Open Annotation Community Group is to work towards a common, RDF-based, specification for annotating digital resources. The effort will start by working towards a reconciliation of two proposals that have emerged over the past two years: the Annotation Ontology [1] and the Open Annotation Model [2]. Initially, editors of these proposals will closely collaborate to devise a common draft specification that addresses requirements and use cases that were identified in the course of their respective efforts. The goal is to make this draft available for public feedback and experimentation in the second quarter of 2012. The final deliverable of the Open Annotation Community Group will be a specification, published under an appropriate open license, that is informed by the existing proposals, the common draft specification, and the community feedback." Below this text are two footnotes: [1] <http://code.google.com/p/annotation-ontology/> and [2] <http://www.openannotation.org/spec/beta/>. A "Reports" section is visible, containing a table of drafts:

Drafts (licensing info)	
2012-05-04	Core Open Annotation Specification (Community Draft 1)
2012-05-04	Open Annotation Extension Specification (Community Draft 1)

On the right side of the page, there is a "Get involved!" section with a "JOIN THIS GROUP" button and a note: "Anyone may join this Community Group. All participants in this group have signed the W3C Community Contributor License Agreement (CLA)." Below this is a "Participants" section listing Paolo Ciccarese as Chair and Robert Sanderson.

“Cross-domain” group

- ▶ A separate group may act as a source for advice and help in design issues
 - may grow out of the current “public-vocab” group by going way beyond schema.org
 - may be set up as a separate group (not decided yet)

2. Vocabulary hosting

- ▶ W3C offers the community to use its namespace:
`http://www.w3.org/ns/XXX`
- ▶ Developers have to
 - document authors' agreement over the content (reflecting consensus with the community at large)
 - best effort for multilingual documentation
 - if it is an update, document that no widely used term is removed
- ▶ Goal is to provide a web service whereby development and maintenance can happen easily

W3C's persistence policy applies

- ▶ At specific sub-areas of W3C all URI-s are kept in perpetuity (as long as W3C exists...)
- ▶ The content of those URI-s *may* change through versioning
 - the W3C Web Site is a giant CVS repository
 - versions are available automatically
- ▶ There are discussions with MIT (one of the hosts of W3C) on further persistence in the case W3C disappears

3. Vocabulary Directory

- ▶ Collect, maintain, and distribute information about vocabularies
 - any vocabularies, not only those hosted at W3C
- ▶ This may require defining additional metadata terms for vocabularies
- ▶ A software environment should be adopted (or developed) to provide this service
 - it should be as open as possible so that the community could jointly maintain it

All these are plans!

- ▶ There is a discussion going on to flesh out the details:
 - see <http://www.w3.org/2013/04/vocabs>
- ▶ Some of these may be part of a planned WG on the best practices of publishing data on the Web
- ▶ We are looking at existing vocabularies and software
 - LOV is a clear example
- ▶ Any input is welcome!

Thank you for your attention!