2020-12-21 CTWG Meeting Notes

Meeting Date

- 21 Dec 2020

Attendees

- Drummond Reed
- Rieks Joosten
- Daniel Hardman
- Brian Dill
- RJ Reiser

Main Goal of this Meeting:

To update CTWG members on the tooling and process discussed at the last meeting and agree on the workplan going forward in January.

Agenda

Time	Item	Lead	Notes
1 min	Welcome & Antitrust Policy Notice	Chairs	
2 mins	Introduction of new members	Chairs	
1 min	Agenda review	Chairs	
30 mins	Report on tooling and process planning meetings	Daniel Hardman Dan Gisolfi	
20 mins	Discuss workplan going forward in Q1	All	
2 mins	Review of Decisions and Action Items	Chairs	
1 min	Next meeting	Chairs	

Recording

- *No recording was made of this meeting*

Presentation(s)
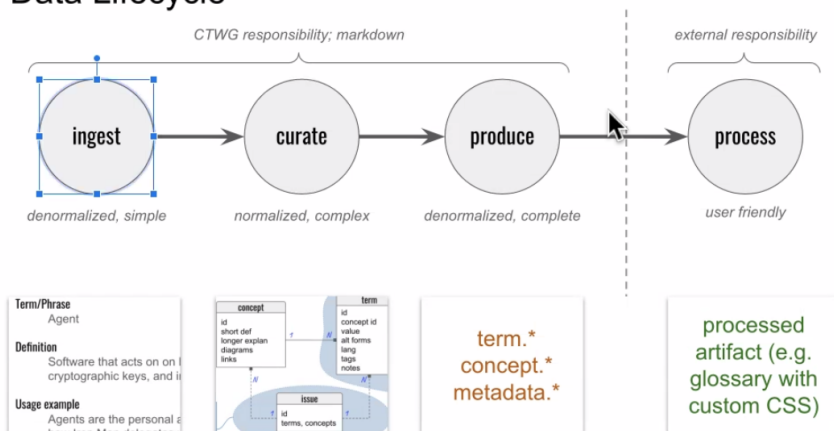
- Daniel Hardman presented this slide deck

Notes

1. New members
2. White papers and using hover links that produce a pop-up
   a. This works the same way as Wikipedia - rich text and graphics but still constrained
   b. All of those in attendance on this call were in favor of doing this
   c. Daniel Hardman pointed out that there may be challenges about what part of the definition shows up in the pop-up
   d. Rieks would like to use a paper we are the authors of to flesh out the specifications—he suggests starting with the proposed Decentralized SSI Governance paper—or one of the Sovrin Governance documents
3. Report on tooling and process planning meetings
   a. Daniel Hardman and Dan Gisolfi met and have some new thoughts
   b. Daniel Hardman and Rieks Joosten met last week and ended out refining the thinking about process
   c. Daniel Hardman shared his takeaways and recommendations in this slide deck
   d. See the Data Lifecycle slide below for the process overview
   e. He showed an example of the ingested form of the data (a proposed definition of the term "Agent"). The ingested form is not normalized.
   f. In the second stage, the data is normalized using a script (that should be fairly easy)
      i. It separates terms from concepts
      ii. Terms are cross-linked to concepts
      iii. This can solve many issues, including multi-lingual terms, multiple terms for the same concept, etc.
   g. We discussed issues around multiple terms for the same concept
      i. In the same language
      ii. In different languages
      iii. Rieks pointed out that all definitions should be provided within contexts (also called scopes or vocabularies)
   h. In the third stage, we "glue them all back together", but we will include metadata that explains what the CTWG knows about the term
      i. Edit history
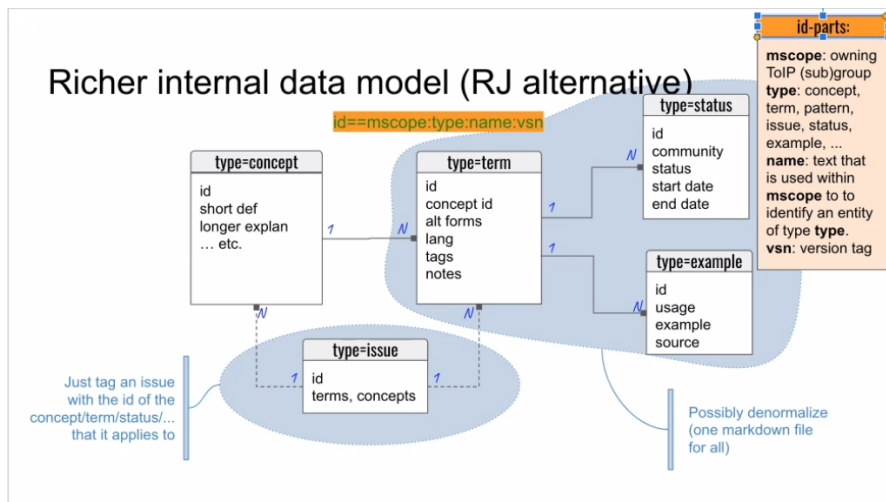   ii. Term status
      iii. Comments

        i. Daniel wanted to see if we had overall consensus on the three forms of data (ingest, curate, produce) as show in the Data Lifecycle diagram below. YES.
       j. Rieks then reported on a richer internal data model (see second slide below)
          i. This allows extensibility of the internal data model to meet more needs over time
          ii. The id of a type=term can be a text string
          iii. This should future-proof the model
       k. Daniel noted that the hyperlinks to the corpus are important. There are two kinds of hyperlinks that are needed:
          i. One is an absolute reference to the corpus as a whole
          ii. The other is a cross-reference within a specific output document
          iii. Rieks made the distinction between links between terms and concepts in vocabularies and in specific documents
          iv. We agreed that hyperlinks to vocabularies and terms within vocabularies
          v. We discussed the actual structure of the hyperlinks and using git artifacts and metadata for the links
          vi. One particular question is version identifiers: do we need human-friendly version identifiers OR git commit hashes OR both?
             1. We discussed these two options and their respective advantages
             2. The git commit hash is "cheap" but not human-friendly.
             3. Rieks suggested that git commits can also have tags.
             4. Daniel liked the idea of tags for the human-friendly portion.
             5. Daniel raised the issue of forks and branches—which is why git uses commit hashes.
             6. Rieks noted that their are indeed good use cases for forking a terminology repo, including personal glossaries, or handing off authority for a terminology.

4. Workplan going forward
   a. We did not have time for this topic
5. Review of Decisions and Action Items
   a. See below
6. Next meeting
   a. Regular time on 4 January 2021

Slides

Decisions

☑   We will follow the three stages of data (ingest, curate, produce) shown in the Data Lifecycle diagram above.

Action Items

- [x] Rieks Joosten will propose the URL syntax for hyperlinking into the CTWG corpus by 04 Jan 2021
- [ ] Daniel Hardman will draft the script for processing ingest entries into the normalize form for curation by 04 Jan 2021