

NIST Semantic Units Repository Project Meeting Monday August 14th – Tuesday August 15th

Attendees

- Stuart Chalk (UNF)
- Bob Hanisch (NIST ODI)
- Ralph Hodgson (QUDT)
- Steve Ray (QUDT)
- Gretchen Greene (NIST DSG)
- Kim Tryka (NIST DSPG)
- Peter Linstrom (NIST DSG)
- Karen Olsen (NIST ASG)
- Gary Berg-Cross (Consultant)
- Evan Wallace (NSIT PEG)
- Adam Morey (NIST DSG)
- Zachary Trautt (NIST MESDG)
- John-Henry Scott (NIST MMSD)

Project Overview

Stuart Chalk introduced himself, gave some background on his experience and current projects and then went through slides detailing his perspective on the project. One question that came up was why pick QUDT? Stuart indicated it is the most mature of all the metrology representation systems that is available.

Steve Ray gave a short presentation on his background on how he came to work on ontologies and linked data. He got involved with QUDT because he had concluded it was the most comprehensive model of units.

Ralph Hodgson presented the history of QUDT from its initial development inside NASA as a funded project to support the NASA Exploration Initiatives Ontology Models (NexIOM) project (part of the Constellation program), to today where it is supported by a not-for-profit organization called QUDT.org. For both Ralph and Steve the project is a labor of love these days. QUDT's 80+ ontologies are taken directly from the VIM and they have also incorporated all parts NIST SP811 as 'things' in the ontology so that they can be referenced.

From Ralph's timeline QUDT has been adopted by Siemens, Allotrope, and OpenPHACTS and is a W3C member. He is also talking to folk at Schema.org who are interested in QUDT being a reference for units added to web pages via RDFa. The representation of units in many domains is not complete because of how much time it takes to generate the units in turtle and checking for conformity.

UnitsML

Karen Olsen gave a report on the NIST UnitsML project that was funded by NIST from 2003-2011. Initiated by Barry Taylor, Frank Olken, Peter Murray-Rust, et al. and funded initially by the Physics Lab, UnitsML is an XML representation of units of measure, and implemented all of what is described in NIST SP 811 i.e.

- SI Base Units
- Derived units included, particularly for named units (Joule, Pascal, Newton, etc.)
- About 50 "root units" are defined in total
- Includes non-SI units such as minutes of time
- Conversion factors (with accuracy)
- Support for aliases and counted items
- Representation of dimensions and quantities

Karen indicated UnitsML tags all units by category, and references them by both name and symbol. In addition, a standard format identifier “license plate” is assigned consisting of the base unit id’s, prefix ids, exponent representation, and indicators of numerator/denominator. Karen also commented there was interest in a “UnitsML Lite” during the development to cover simple use cases.

An OASIS Technical Committee was formed in 2006 to move forward the UnitsML as a formal standard. OASIS was chosen as it was perceived as a faster mechanism to get UnitsML as a standard in ISO. The TC was active until late 2011 and was formally closed by OASIS in 2016.

Sadly, the work at NIST lost funding in 2011 and was not able to keep momentum going after that. The OASIS TC also ran into problems getting the standard ready for balloting due to the OASIS requirements of i) two reference implementations of the standard and ii) the organizational composition of the TC. Therefore, currently UnitsML is still only an internal NIST project even though it is ‘nominally’ complete.

Ralph commented that he should be able to quickly convert UnitsML schema into an ontology (which he did within 45 minutes) and that he would do a comparison of QUDT and UnitsML to see if they are semantically consistent. Zach asked why would someone choose QUDT over UnitsML and the answer was because UnitsML is not an official NIST product.

Project scope and implementation: Identifying issues and opportunities

There was a general discussion about the project scope and implementations. The discussion culminated in Bob proposing that a product that could be hosted at NIST was a units repository for interoperability. This would be unit representation agnostic and serve the community by allowing users to not only find units but translate between unit representations. The rest of the participants were in enthusiastic agreement about this.

The rest of the discussion revolved around what was out of scope (below) and what activities to prioritize. Bob suggests we focus on 6-8 use cases in areas such as toxicology, forensics, materials, additive manufacturing, energy, greenhouse gases as they align with NIST activities and thus would provide opportunities to demonstrate the importance of this activity to other groups within NIST. This will be done in time for the next project meeting at NIST (February-March 2018). In addition, we should reach out to other organizations, e.g. Allotrope (Don Rolph), OpenPHACTS, CSIRO, W3 and Schema.org, to get use cases from them that would be addressed after the next project meeting.

Ideas that came up that are out of scope for year 1:

- converting numeric values (providing a repository API service to convert a numeric value from one unit to another)
- currency (will be important for commerce though)
- don’t worry about ‘complete’ unit coverage, focus on only one/two domains
- unit recommender service
- internationalization – providing a multilingual interface to the repository

Ralph encouraged us to keep in mind that we need to work toward a minimum viable product in the initial stages of the project, especially thinking about desirability and feasibility.

Creating a Metrology Ontology

Stuart suggested that the VIM should be directly translated to a metrology ontology as it seemed appropriate to have a semantic equivalent to the VIM. Ralph indicated that this was what they had already done with QUDT and that they had also integrated SP 811 into the ontologies. This was done using the third edition of the VIM and so Ralph agreed to look at the 2012 revision of the third edition to make sure QUDT was up to date.

Semantic Units Repository – Workflow, functionality & implementation

Stuart presented a few slides showing the ideas he put in the proposal about the design of the repository. Gary had some questions about the workflow relative to the developer part of the website and should it not be linked to the NIST review side. Stuart indicated that he thought it not necessary as the interface of the developer site would be designed so that (as best as possible) users were constrained in the creation of a new unit representation to metrologically appropriate choices. Still this is an important point and will be integrated into the workflow analysis.

Peter also indicated he had misinterpreted the way in which the deployment of the repository was to be implemented. Peter thought that once the repository was complete domain groups could download it and run a local version to store units for their community, with a copy at NIST aggregating the content of the remote versions of the repository – a federated approach. Stuart indicated that he had envisioned it being a signal instance at NIST, with replication of that instance at other NMIs. Others were supportive of the federated approach although the issue of which remote repositories do you aggregate (maybe some are inappropriate) is a touchy one. Further discussion is needed on this issue as no firm decision was made at the meeting.

Workflow

Aside from the comments above related to the diagram Stuart presented, there were not many additional comments except that the workflow should be evolved as we go through the process of looking at the use cases and addressing the needs that come up. It was also suggested that we organize a beta testing group and address issues that come up then.

Functionality

There was a good discussion about functionality from the perspective of:

- a) functionality of the repository interface
 - Provisioning /SPARQL (search)
 - Develop a Unit Creator
 - Does my unit exist in UCUM?
 - Create this unit please! (request)
 - This unit is incomplete - please help (like Wikipedia)
- b) services of the repository
 - REST API – lookup service (units, quantity kinds, property), crosswalk service, etc.
 - Does my unit exist in another unit system?
 - Downloads of units of a particular type by system, quantity kind, etc.
- c) ones in support of the initial development of the repository
 - Units crawler for DBpedia (what units are represented?)
 - Check existing QUDT Units on GitHub
 - Create QName generator (link to unit if it exists)

Of course functionality will be added as it is needed going through the use cases.

Implementation

Stuart indicated that he would use Apache, MySQL, PHP, Fedora Commons, and Bootstrap (UI) to develop the repository. When asked but what she thought would be the best implementation from a NIST perspective, Gretchen indicated it did not matter as long as you could appropriate create the services needed. Stuart suggested that the site should use two-factor authentication for user access (for unit developers and reviewers) and Adam indicated that would not be a problem if we wanted to use it.

Outreach: Identifying participants, stakeholders, and communities of interest

The group had a good discussion of the groups that we needed connect with this project. Ralph suggested that we create a one-page charter for the project that could be used when contacting these groups (Stuart agreed to create this). We need to think about how to liaise with external groups and what level of involvement do we need?

- Stuart - Publishers, CSIRO, Unit Ontology, UCUM - Clem McDonald, RDA, WikiData
- Steve - W3C, NASA, QUDV, ASHRAE
- Ralph - Allotrope, Schema.org, OpenPHACTS, TNO
- Bob - NMI directors, CODATA/ICSU, IVOA

Project communication tools

Stuart suggested using Atlassian confluence as the main area to admin the project, however it was suggested to start by using the Wiki and Task features of GitHub and see if that will work. Therefore the tools to be used will be:

- Google Docs (Advisory Board feedback/input mechanism)
- Google Group (for Project team/Advisory Board Discussions)
- GitHub - Unit Repository code, QUDT code/files

Action Items

In addition to the outreach outlined above the following action items were proposed:

- Discuss how we will get QUDT content into MySQL (Stuart, Ralph, Steve)
- Get MML strategic plan (Bob)
- Complete survey of unit ontologies (Stuart, Steve, Ralph)
- Create 1 page project charter (Stuart) (example?)
 - Talk about: the federated repository in a box; Enabling interoperability; Supporting FAIR, Reproducibility; Discoverability; Historic/important activity; Building upon a significant amount of groundwork; Translation of the major work in metrology to the digital space
- Send out news item to RDA list, advertise BoF (Stuart)
- Advertising – put something on the UnitsML page? (Karen)
- Project page for NIST – Send to Gretchen (Stuart)
- Put on W3C website (via NIST?) (Steve?)
- UnitsML harmonization (Steve/Ralph)
- Name this project with a short title and create a logo (Stuart)
- Review other open source software (i.e. wikidata)
- Could MDCS be used as a MVP? (Bob)
- Cordra by CNRI (LaTeX/mathjax support?)
- Viewer from TopBraid, TopBraid EDG can be put up for authoring (Ralph)
- Look at the difference between SHACL vs OWL (<http://spinrdf.org/shacl-and-owl.html>)