

NIST Semantic Units Repository Project Meeting 2 Friday April 6th 2018

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)
- Evan Wallace (NSIT PEG)
- Adam Morey (NIST DSG)
- Zachary Trautt (NIST MESDG)
- John-Henry Scott (NIST MMSD)
- Ray Plante (NIST DSG)

Slides from the meeting are available at <https://units.unf.edu/project-meetings>

Introduction, review of charter, progress to date

Stuart provided a short review of what was decided at the first meeting last year about the Units of Measure Interoperability Service (UMIS), in particular the shift in focus of the project to create an interoperability website for units that is not aligned with any unit representation system (required by NIST). Stuart then indicated the progress on the project, i.e.:

- Survey of unit representations
- Development of an alpha version of the website
- Initial API endpoints (not formalized)
- Added functionality/features based on two use cases
- Explored FAIR and how UMIS can support it

Stuart then discussed the FAIR principles and highlighted some that need thought in the project, i.e.:

- Accessible: Metadata are accessible, even after the data are no longer available
- Interoperable: Vocabularies (ontologies) that follow FAIR principles
- Reusable: Metadata are released with a clear and accessible usage license

Stuart also presented slides on the current focal point of FAIR – Global Open (GO) FAIR. This activity is part of the European Union’s focus on open data – through the Horizon 2020 project. Briefly, the project is about creating ‘Implementation Networks’ to support societal change (GO CHANGE), provide training (GO TRAIN), and build the technical infrastructure for the Internet of FAIR data and services (IFDS). An analysis of FAIR in the context of UMIS is important and needs to be articulated so that can be part of the continued development of UMIS.

After the meeting thoughts:

1. In discussion with Steve Ray and Ralph Hodgson it was realized that creating a page that provides the representations of a particular unit of measure should not be given a persistent identifier as we don’t want the community to use such an interoperability page as something to reference to. Similarly, if a representation system has a website we don’t want the UMIS representation system page to become the defacto URL for that representation system.
2. Within the context of FAIR the UMIS website might be one of the pieces of technical infrastructure that a Metrology Implementation Network (MIN) would oversee for the Internet

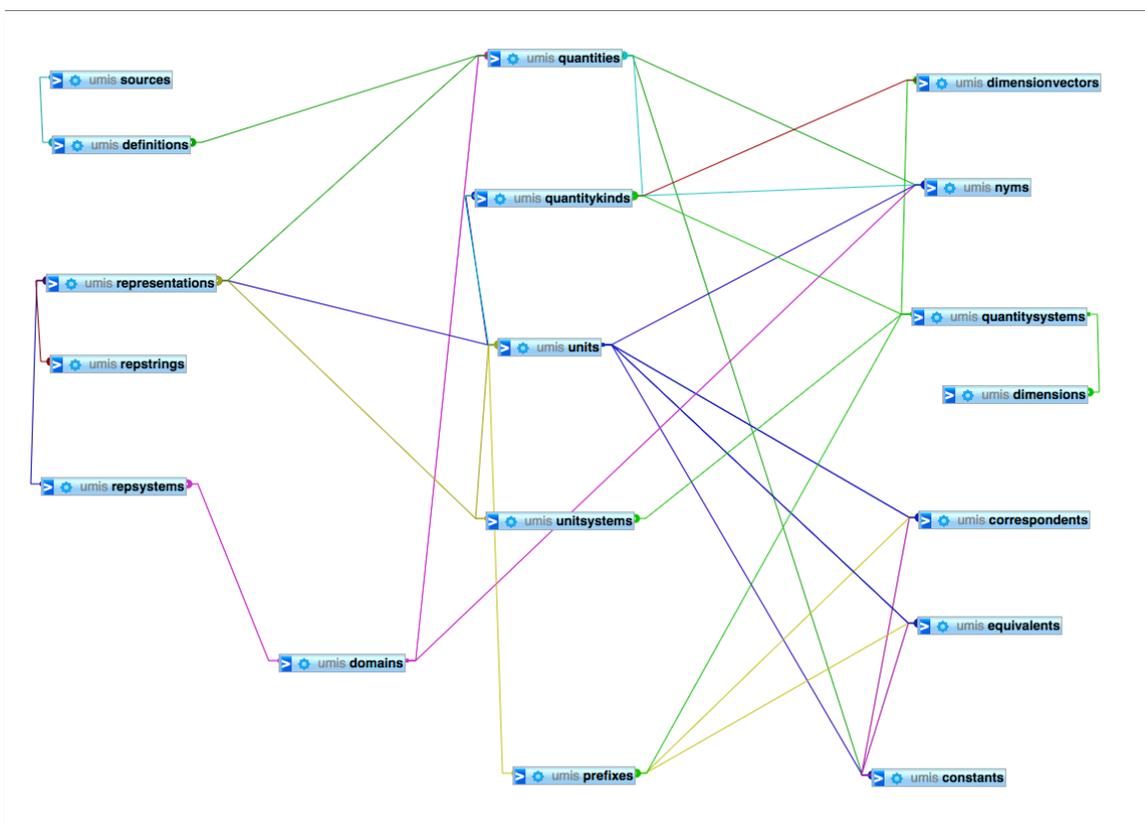
of FAIR Data and Services. Stuart will discuss this with participants of the NPL meeting in the UK (May 1 – 3) which includes Barend Mons the head of GO FAIR. If there is to be a MIN then it would logically be instantiated in the NMI's (and the BIPM of course).

Overview of the alpha version of the UMIS portal

Stuart then described the alpha implementation of the UMIS portal (<https://untisunf.edu/umis/>) where the focus had been more on functionality than on content. As of the meeting the following endpoints were available:

- Units – index, view, crosswalk, add (search)
- Unit Systems – index, view
- Unit Representation Systems – index, view
- Unit Representations – index, view (add)
- Quantity Systems – index, view
- Quantity Kinds – index, view
- Quantities – index, view
- Dimension vectors – index, view

Stuart noted that the current schema (below) has a lot duplication for the representations as each representation is a different entry in the database even if it is the same string is used to represent a unit. This will be refactored in the next iteration of the schema so that each unique representation of a unit is only present once. Gretchen asked if a model had been created for UMIS and Stuart indicated that it had not. It was generally agreed that this should be done to aid in optimization of the schema.



Database schema for the alpha version of the UMIS portal (not all tables are used yet)

Stuart indicated there were many features that still needed to be addressed including:

- *Representation systems* – initially the scope of representation systems was limited to formal encoding standards like QUDT and UCUM, however going through the development of the portal it became clear the scope could/should be expanded to include the formal guidelines of best practice for specific communities (i.e. IUPAC for Chemistry, IUPAP for Physics etc.), and more general guidelines (e.g. from NIST SP-811).
- *Representation adoption* – also in certain situations organizations/groups/communities have adopted a particular representation system (i.e. HL7 has adopted UCUM). How to represent this efficiently? See below...
- *Adding format variants* – for each representation string there are multiple ways they can be formatted, e.g. plain text, HTML, and LaTeX. These should be added so that these variants of a representation string can be found/used by users of the site and also available as encodings in the download format.
- *REST API* – although a few of the endpoints have JSON download options the API has not yet been built. This will be completed over the summer and documented with using Swagger

One feature that came up was the idea of a user having a specific perspective when using the site. This could be implemented with/without have user accounts created by asking a new user what discipline they are from and remembering that in a cookie.

At one point in the discussion Peter asked units encoding system was being used to represent the units. Stuart acknowledged was not using a units encoding system at this point. After some discussion it was decided that the VOUnits encoding specification would be used (see <https://arxiv.org/pdf/1509.07267.pdf>)

There was also discussion of the best way to present the different representations of a particular unit to a user. Classification of representations could be useful, i.e. ‘preferred’ representation (which could be different domains/disciplines), acceptable, unacceptable, or wrong. Show only the ‘preferred’ representation? Show representations by frequency of use across different representation systems? Will need to look at the right way to advise user on what they should use based on their situation.

Other topics that came up but have not yet been addressed were:

- How to balance findable/easily discoverable with “curated” and managed units (curation takes a lot of effort and would slow down new additions to the UMIS site)
- How are the units etc. the basis for SI traceability? (out of project scope)
- How do we deal with provenance?

Topics considered out of scope for this year:

- Physical constants and conversion factors
- Internationalization (get the NMI’s to help?) – how to deal with Unicode variants of characters?
- User profiles

Out of scope of the project:

- Services for integration into other websites/tools (e.g. Excel)

Use cases addressed and proposed solutions

Solutions to two use cases provided by Peter Linstrom where implemented.

Use Case 1: A researcher is preparing to publish supplemental data for a research article. As is often the case, the researcher will supply the data as an Excel file. How can the researcher use UMIS to make it easy for others to unambiguously identify the units in the file?

Solution: Create a permanent link to the unit representation used (Stuart assumed that all units in the Excel file are from the same units representation system)

As an example (see slide 20 of the presentation) a permanent link to <https://units.unf.edu/umis/repsystems/view/IUPAC> could be used to indicate that the author had used unit representations consistent with the IUPAC guidelines for units from the IUPAC Green Book.

Use Case 2: A researcher has data with units identified by a particular coding system and wants to merge this data with data with units identified by the Unified Code for Units of Measure. How can the researcher match up units from the two systems?

Solution: Create a units cross-walking service

Stuart showed the functionality at <https://units.unf.edu/umis/units/crosswalk> to the group (slide 22 in the presentation) when two unit representation systems can be selected and the equivalences shown and if required downloaded as a JSON file.

Revisiting an ontology of the VIM. Creating a UMIS ontology

Stuart again brought up the topic of the need for an ontology representation of the VIM. The original thought on this at the last meeting was that there did not need to be one as QUDT already has it in their core ontology. However, in terms of providing an API for access to the content of the UMIS portal Stuart felt that an appropriate ontological representation of the different pieces of the metrology space would be important. There is also a need to add some ontology terms for the structure of the data returned for the UMIS (i.e. the predicate 'hasEquivalentUnit' is not part of the VIM). After discussing this issue in the meeting (and after Stuart will create JSON-LD output for units and use QUDT terms and others to represent the semantics).

Technical discussion – database schema

There was not too much discussion about the database except what had come up before. Need to think about adding administrative functions – web-based updating and maintenance, site usage statistics (Google Analytics). Once the site goes live there will be other admin functions to add but those are down the road.

Brainstorming system needs/functionality – quantities, quantity kinds, dimensions

Issues around domain-based usage (adoption) of different unit representation systems by specific organizations was discussed and it was decided to handle these in documentation on the website. Additionally, it was agreed that statements about common usage of units was needed as well as comments about common mistakes in unit representation. Stuart also brought up the issue of unit representation in software (e.g. Astropy - <http://www.astropy.org/>) and it was decided this was out of scope.

Deliverables, priorities, outreach

There was no additional discussion of the deliverables of the project outside of what is in the charter. Stuart will work with Bob to make sure that the site lives up to NIST's expectations.

As for outreach, the internationalization of the website was deemed out of scope at this point but is certainly something to bear in mind moving forward. Stuart indicated that he had an idea to put together an Excel spreadsheet that organizations could use to transit their units usage back to the project. It was pointed out that this does not scale well...so it will be left as an open idea right now. The idea of web scraping for units came up and Stuart will explore getting units different sources that are convenient to scrape.

Additionally, many sources of use cases were suggested, including:

- The Interagency Working Group on Open Science (IWGOS) (disbanded?)
- Networking and Information Technology Research and Development (NITRD) Program (<https://www.nitrd.gov/>)
- CENDI – Interagency Working Group (<https://www.cendi.gov/>)
- Data service providers
- Office of Scientific and Technical Information (OSTI) (<https://www.osti.gov/>) – Carly Robinson
- Jane Greenberg at Drexel (Dryad - <https://datadryad.org>), Ian Fore, Vivien Bonazzi (NIH), RDA (Rebecca Koskela, Keith Jeffery), CODATA (Phil Archer, Simon Cox)
- Center for Data to Health (CD2H) – NIH Melissa Handael

Action items (Stuart)

- Develop a model for the UMIS system
- Update the schema for UMIS to pull apart the representation strings from the representations to avoid duplication and make it easier to create HTML and LaTeX version of each one.
- Review FAIR principles and develop white paper on how UMIS will be implemented in compliance with these principles.
- Add in lots more data about representations (start with SP-811, ISO 80000, and IUPAC Green Book)
- Implement the VOUnits encoding specification in UMIS portal
- Build out the API, document and expose using Swagger