Development of a Units Repository for Administration and Dissemination of Scientific Units for Use in Semantic Applications

Stuart Chalk, Department of Chemistry, University of North Florida
Ralph Hodgson, TopQuadrant and QUDT.org
Steve Ray, Carnegie Mellon University and QUDT.org

Submission to the NIST **Material Measurement Laboratory (MML) Grant Program**

Summary

There are a number of disparate activities currently focused around the digital representation of metrology concepts because of the urgent need for a way to definitively represent and refer to scientific units, quantities and dimensions. Interoperability is significantly hampered in the current, fragmented digital unit landscape.

This project is therefore focused on leveraging the work on QUDT, UnitsML (a NIST project) and NIST's expertise in metrology to develop a web portal for the administration, documentation and usage of semantically defined scientific units. The goals of this project are to:

- Provide a stable, authoritative source for scientific units *for digital applications*
- Allow international units-related bodies to administer the process of *digital* unit development
- Allow subject domain experts to propose new units, quantities, scales and conversion factors etc.
- Codify, through ontologies, vocabularies and naming conventions unit and quantity representations
- Promote systematic unit/quantity application and usage through best practices and use cases
- Allow international standards agencies to provide formal language translations of units
- Provide a mechanism whereby legacy units can be represented and related to current units
- Produce a global network of synchronized unit repositories

Rationale

The introduction of the Semantic Web (1) and more recently the move toward FAIR (Findable, Accessible, Interoperable, and Reusable) (2) data has mandated a new perspective on the representation of scientific data and metadata. Where previously human understanding and interpretation of the way in which scientific data were reported was only semi-standardized, translation to the digital domain requires a much more rigorous treatment of both meaning and representation, for both data and its units. As the inventor of the World Wide Web, Sir Tim Berners-Lee wrote, "*The established system of journals for communicating the results of scientific research is already challenged by the existence of the web. But we are only at the early day of a new Internet revolution, one which will have a deeper and more disruptive impact on scientific, and other, web publishing, and have profound implications for the web itself. An emerging successor for the web, the Semantic Web, will likely profoundly change the very nature of how scientific knowledge is produced and shared, in ways that we can now barely imagine*." (3)

Scientific unit definition, representation, and usage are of huge importance in the move to open and interoperable science and have not been prioritized globally in the digital realm. Now is the time to address the needs of the scientific community with a comprehensive system for representation, administration, and dissemination of scientific units through development of metrology standards and a

unit repository (web application) that supports both human and computer needs for usage of digital scientific units.

This project proposes to bring together approaches, formats, implementations and lessons learned from current projects in the semantic representation of units with the expertise of NIST in the area of metrology. While the development of existing projects have been fuelled by necessity (application/domain driven), this project seeks to develop a high level (both international and scientific) formalization of concepts in metrology (i.e. as a metrology ontology) and development of an online system to disseminate digital scientific units, administer the creation/updating of said units and quantities, and describe their usage in a standardized form. Such as system would allow the community to develop implementations that fit their need, yet conform to a system that ensures the interoperability and reusability components of the FAIR perspective on scientific data.

A significant issue, that this project would address, is the development of a complete set of scientific units – i.e. for every discipline – across all unit systems in current usage. This has been a major problem for every current project in this area – it is impractical for small groups to achieve – and is logically outside the scope of each project. A much better approach, fundamental to this project, is to build a framework from which any scientific community can create units, articulate equivalencies, describe application areas, define terms (domain specific) and numeric bounds, and articulate and share those with the rest of the scientific community. It is also important to state here that there is no intention (or need) in this project to redefine terms in metrology as i) they are already well defined and ii) only a digital representation of those concepts is the aim.

Building a Metrology Ontology

The initial phase (six months) of this project will focus on the analysis of metrology concepts and the translation of those into a formal metrology ontology (MO). This will be accomplished by bringing together expertise from the NIST metrology group, the developers of the Quantities, Units, Dimensions, and data Types (QUDT) project (4), and the work on the Units Markup Language (5). The work will focus on accurately capturing metrology concepts in a semantic representation, i.e. by defining classes, properties, domain and range. In addition, the development of the MO will consider the non-semantic application of units and quantities – meaning the representation of scientific units in non-semantic formats. This is critical to encouraging the migration of scientific data to the semantic realm.

About Quantities, Units, Dimensions and Data Type Ontologies – QUDT

This proposal builds off the QUDT project and leverages the existing community of scientists that are already using it in their applications. The development of QUDT was done as part of the NExIOM program at NASA in the late 2000s. The ontology suite was part of a larger semantic suite intended to represent data in the Constellation program before ultimately that program was cancelled. QUDT has lived within NASA since that time and version 1.1 was openly published on qudt.org in 2012. Currently, version 2.0 of the ontology suite is in its final development, focused on solving a number of issues related to version 1.1, notably adoption of well-known names for prefix Qualified Names (QNames) and improved ability to dereference Universal Resource Identifiers (URIs) using W3C Semantic Web recommended practices.

A quick search of Google will return ~59000 hits on QUDT the acronym – an indication of its status as the most implemented digital representation of metrology related concepts to date. It is by far the most

organized, comprehensive, and thus usable ontology for quantities, units, and dimensions for general scientific usage. It provides definition of unit classes based on quantities as well as application areas, includes a complete set of CODATA defined physical constants, over 200 defined quantities, nearly 1000 defined units, and nearly 500 defined dimensions, as well as RDF transformation of ISO 80000 parts 1 to 14. QUDT also includes definition of metadata terms important in the description of facets of quantities, units, and dimensions and addresses units in multiple unit systems.

Development of QUDT is now coordinated through QUDT.org. QUDT is a 501(c)(3) not-for-profit organization founded to provide semantic specifications for units of measure, quantity kind, dimensions and data types. QUDT is an advocate for the development and implementation of standards to quantify data expressed in RDF and JSON. Its mission is to improve interoperability of data and the specification of information structures through industry standards for Units of Measure, Quantity Kinds, Dimensions and Data Types. QUDT.org is a member of W3C and is actively engaging with the community to promote its mission of "*quantified data in the web*".

About the Units Markup Language (UnitsML)

The UnitsML project (5) was initiated at NIST in 2006 as a response to customer need to represent scientific units in the Extensible Markup Language (XML) and to consolidate work on units within other activities at NIST. The intent was to develop a national reference for markup of units but specifically not to define units and quantities. As NIST is 'not in the business of publishing standards' the project committee coordinated development under OASIS (6). From the beginning, the project used the XML 'Meta' approach for definition of units – that is, the formalized description of a unit based on the fundamental base units and prefixes. While more complicated than just identifying a unit in an attribute in XML, this approach allowed accurate definition of units and enhanced their interoperability.

The development of UnitsML continued through the OASIS-TC through the end of 2011. An essentially complete (v1.0) XML schema had been developed at that time (7) and publications from the committee about its usage (8) were in the literature. NIST had internally developed a database of units (UnitsDB) and the committee was ready to initiate the approval process at OASIS. However, due to lack of funding the approval process never got started. Therefore, the work on UnitsML has never been formally standardized and has been implemented in only a handful of applications.

This project intends to take advantage of the significant development effort in UnitsML and the institutional knowledge of its development within NIST. Although this project is focused on the semantic representation of units, it is important to understand the needs of other data formats for unit representation such that anything coming from this project can be applied in XML and other formats. The implementation of units in UnitsML also provides a different viewpoint to other projects on units that is highly valuable.

Other Digital Implementations of Unit Systems

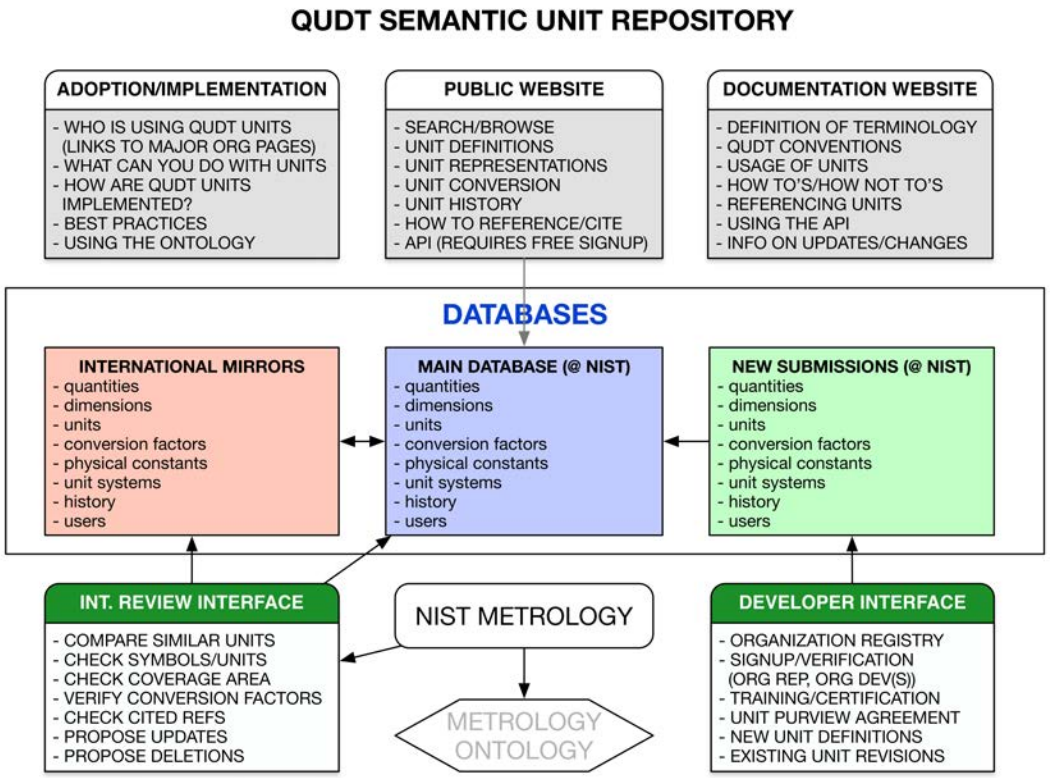
The outputs of other projects that have developed unit representations, and projects that are about to start, will be also be considered in the process of development of the MO, notably:

- Units of Measurement Ontology (UO) – part of the OBO foundry (9)
- Quantities and Units of Measure Ontology Standard (QUOMOS) (10)
- Units in Semantic Web for Earth and Environmental Technology (SWEET) (11)

- Units in the Analytical Information Markup Language (AnIML) (12)
- Units in the Geographic Markup Language (GML) (13)
- Units in MathML (14)
- Units in the Scientific, Technical, and Medical Publishing (STTML) - part of CML (15)
- VOUnits - International Virtual Observatory Alliance (IVOA) (16)
- Unified Code for Units of Measure (UCUM) (17)
- Schema.org (18) is in discussion with QUDT.org to provide mappings of Resource Description Framework for Attributes (RDFa) based representations of units to QUDT ontologies. The interest arises in the group that is working on "Internet Of Things (IOT)" (19).

Building a Units Repository

The second phase of this project will focus on the development of a unit repository for administration and dissemination of units. The figure below shows the initial conceptualization of the system.



Key features of the units repository are:

Public website – search and browse capabilities for units based on free text and/or unit system, discipline area, quantity. Additionally, the API will be documented (Swagger) with examples for usage. Free user accounts will allow API key generation (for usage) and alert scheduling for updates.

Documentation and adoption/implementation website – complete documentation will be provided relative to the MO terms, ontology conventions, unit citation, unit versions, and how to use the API. Separately, a website documenting the current adoption of QUDT units, examples, implementations, best practices, will be developed and maintained.

Databases – the main and new submission databases will be replicated at each mirror site of the units repository. Updates to these databases will be automatically broadcast to the other mirror sites the system and those changes integrated after change collision analysis is performed.

Developer website – this is for representatives and developers of Domain Unit Authorities (DUA's) to register and administer units for their organization/domain. This interface allows the addition of new units and updates to existing units to be submitted to the new submissions database – a database that it separate from the main database so that inclusion of new units can only be authorized by the NIST metrology group after internal review. This site will also be used to train and certify developers so that technical issues can be mitigated and the NIST review process can be as efficient as possible.

Internal review interface – this is for the NIST metrology group to be able to run consistency checks, i.e. check units of the same scale, unit representations, conversion factors, and propose updates and changes to the existing units database. The website would a mechanism to restore changes should a decision be made in error and/or a revision subsequently be reversed.

Other Activities

During the first six months of the project (concurrently with the ontology development) international organizations responsible for units will be contacted, brought up to speed on the project and surveyed for needs of the unit repository (this will be considered the projects' Advisory Committee). This will allow time for development of the functionality required to administer units with the repository including (but not limited to); adding new units, updating existing units, deprecating (or reclassifying as legacy) units, defining legacy units, internationalization, physical constant representation, and conversion factor representation. The group will also be tasked with aggregating a list of additional international organizations (e.g. the International Unions) that will be defined as a 'Discipline Unit Authority' (DUA) that will be given ultimate authority to administer units for their scientific field (e.g. the International Union of Pure and Applied Chemistry – IUPAC – would be the DUA for Chemistry).

A website (not the unit repository) will be developed at the start of the project to serve as a central place for information about the project, including progress reports, meeting attendance, presentations etc. A white paper on the project will be developed (first two months) and made available on the project website to report to the community the purview of the project, the timeline and how appropriate entities can get involved. Finally, the developers will co-ordinate with NIST representatives to communicate with other National Measurement Institutes around the world about hosting repository replication sites (including software).

Deliverables

D1: Ontology Suite – an ontology-based framework foundation for the description of the following:

a) Concepts in metrology
b) SI Unit System, prefixes, and derived SI units
c) Quantities
d) Dimensions
e) Conversion factors
f) Fundamental scientific constants
g) Naming and identifier rules
h) Core collection of data types

D2: Reference Implementation – An implementation of the Ontology Suite for the base SI units system

D3: Semantic Units Project Documentation – including usage examples in different formats, use cases, best practices

D4: Units Repository – a web accessible, governed site that includes a units database (UnitsDB) hosted at NIST

D5: Application Programming Interface (API) for the Units Repository

D6: Validation service (also hosted at NIST) for implementers to verify they are:

a) Referencing units correctly
b) Representing units correctly
c) Converting units correctly

D7: Assignment of Digital Object Identifiers (DOIs) – a report assessing DOIs as appropriate to parts of the above framework

Proposed Schedule of Activities (August 2017 – July 2019)

The proposed project timeline is shown below. The timeline is aggressive because i) the need for digital units is urgent and ii) the impending redefinition of the SI units presents an opportunity to focus significant attention on the need for digital units.

- Pre-announce project to key organizations to find representatives for Advisory Committee (AC) (Jun 2017)
- Develop a website for communication of activities and submission of requests, suggestions, and comments (Aug 1 2017)
- Formal announcement of project, call for input (Aug 1, 2017)
    - Communicate with contacts to find representatives of additional organizations
    - Encourage submission of use cases and current issues revolving around digital units
- Formation of the Advisory Committee (AC) (early Aug 2017)
- First meeting of Working Group (WG) with remote AC participation (mid Aug 2017)
    - Discussion and development of approach with goals
    - Identification of example use cases
    - Develop outline for project development
- Development of the Metrology Ontology (Sep 2017 – Feb 2018)
- AC Meetings to conduct needs assessment for unit administration (Sep 2017 – Feb 2018)
- Collection of resources (Sep – Oct 2017)
    - Resources describing activities to develop digital representations of QUD's
    - Standards in metrology, weights and measures, fundamental physical constants
- Publish first WG meeting report (Sep 15, 2017)
- Publish project white paper (early Oct 2017)
- Meetings to coordinate remote hosting of unit repository replication sites (Oct – Nov 2017)
- Publish Metrology Ontology report and updates to QUDT based on the report (late Feb 2018)
- Second meeting of WG with remote AC participation (mid Mar 2018)
    - Review of Metrology Ontology
    - Review AD needs analysis
    - Define Units Repository architecture, development timeline, testing, progress metrics
- Publish second WG meeting report (Apr 7 2018)

- Develop alpha version of units repository (Apr - Aug 2018)
- Third meeting of WG with remote AC participation (mid Aug 2018)
    - Review of alpha version of units repository
    - Develop and document naming conventions to be used in units repository
    - Define scope/implementation of the SI base unit reference implementation
    - Make decisions about DOI assignment and format
- Publish third WG meeting report (early Sep 2018)
- Develop beta version of units repository (Oct - Feb 2018)
- Start formal documentation of unit repository (Feb 2019)
- Fourth meeting of WG with remote AC participation (mid Mar 2019)
    - Review of beta version of units repository
- Publish fourth WG meeting report (early Apr 2019)
- Develop API documentation (Apr 2019)
- Develop first release version of units repository (Apr - Jun 2019)
- First version release of units repository, metrology ontology and final report (late Jul 2019)

References

1) Semantic Web https://www.w3.org/standards/semanticweb/
2) Find, Access, Interoperate, Re-use (FAIR) Data http://www.datafairport.org/
3) Berners-Lee, T.; Hendler, J., Publishing on the Semantic Web - The Coming Internet Revolution will Profoundly Affect Scientific Information. *Nature* **2001,** *410*, 1023-1024 http://dx.doi.org/10.1038/35074206
4) Quantities, Units, Dimensions, and data Types (QUDT) http://qudt.org
5) Units Markup Language (UnitsML) http://unitsml.nist.gov/
6) OASIS UnitsML Technical Committee https://www.oasis-open.org/committees/unitsml/
7) UnitsML Schema http://unitsml.nist.gov/Schema/UnitsML-v1.0-csd04.xsd
8) http://nvlpubs.nist.gov/nistpubs/jres/115/1/V115.N01.A03.pdf
9) Units of Measurement Ontology https://bioportal.bioontology.org/ontologies/UO
10) Quantities and Units of Measure Ontology Standard (QUOMOS) https://www.oasis-open.org/committees/quomos/
11) Units in Semantic Web for Earth and Environmental Technology (SWEET) https://sweet.jpl.nasa.gov/
12) Units in the Analytical Information Markup Language (AnIML) https://animl.org/
13) Units in the Geographic Markup Language (GML) http://www.opengeospatial.org/standards/gml
14) Units in MathML https://www.w3.org/Math/
15) Units in the Scientific, Technical, and Medical Publishing (STTML) - part of CML http://www.ch.ic.ac.uk/rzepa/codata2/
16) VOUnits - International Virtual Observatory Alliance (IVOA) http://ivoa.net/documents/VOUnits/
17) Unified Code for Units of Measure (UCUM) http://unitsofmeasure.org
18) Schema.org http://schema.org/
19) "Internet of Things" on Schema.org http://iot.schema.org/
20) Atlassian Confluence https://www.atlassian.com/software/confluence
21) Google Docs http://drive.google.com/
22) GitHub https://github.com/