



Semantic Units Repository: Project Overview

Stuart Chalk



August 14th, 2017



Who Am I?



Stuart J. Chalk
Associate Professor of Chemistry
University of North Florida

- Ph.D. Analytical Chemist (21 years @ UNF)
- Moved into chemical informatics (last five years)
 - Extraction of locked chemical property data
 - Scientific data models
 - Vocabularies
 - Knowledge representation
- Skills – MySQL, PHP, JavaScript, JSON-LD, XML, RDF
- Working with – RDA, IUPAC, Springer



What are my Research Interests?



- Projects (Past and Present)
 - AnIML (<http://animl.org>)
 - UnitsML (<http://unitsml.org>)
 - NIST SDS Project
 - SciData (<http://stuchalk.github.io/scidata/>)
 - Open Spectral Database (OSDB – <http://osdb.info>)
 - IUPAC Gold Book (<http://goldbook.iupac.org/>)
 - ChemExtractor (Springer funded)
 - ChemCurator



What is this Project About?

- Creation of digital representations of scientific units for use in computer applications (it is not about defining units)
- Building a repository of unit representations (instances) that is self consistent and accurately describes each unit in terms of its metrology and provenance
- Development of a workflow to allow domain unit authorities (DUA's) to create unit representations for units within their domain using the QUDT ontology suite
- Provide mechanisms (through the repository) to definitively reference to unit representations (e.g. DOI assignment)



Why Did I Submit This Project?

- Move to open data/science and linked data is underway and there is lots of data – units are fundamental to the infrastructure
- Representation of data – FAIR (<http://www.datafairport.org/>) (Findable, Accessible, Interoperable, and Reuseable)
- Need unit services from authoritative source to support research
 - support for machine processing of scientific data (with history)
 - definitive conversion factors, physical constants
 - units lookup, value conversion (with error propagation)
 - PIDs associated with dimensions, quantities, units, constant, etc.
 - support for digital research notebooks



My Perspective on this Project

- Any project on digital units has to be a community effort – no one group has the authority to define the digital version of units and it's impractical
- NIST is the authoritative and trusted organization for units – this project needs its experts in metrology to ensure the appropriate and accurate implementation of metrology concepts
- This is not about creating a standard, more a platform for stakeholders to create consistent representations of units
- Persistent identifiers are very important for computer and human access to definitions and representations of metrology concepts and instances



Deliverables (Year 1 of 2)

- Communication website for activity/event/progress reporting and submission of requests, suggestions, and comments
- Collection of resources on units/quantities/dimension/constants etc.
- White paper describing the project plans and avenues for participation
- Aggregation and publication of unit use cases and discipline specific needs
- Development and publication of a metrology ontology in support of the project
- Update QUDT schema to implement metrology ontology terms/improvements based on use cases and NIST input
- Publication of the web application development plan for the Semantic Unit Repository – technologies, architecture, services, capabilities, security, access and replication
- Development of an ‘alpha’ version of units repository showing proof-of-concept



Project Management Tools

- Atlassian confluence (for project personnel)
- Google Docs (Advisory Board feedback/input mechanism)
- Google Group (for Advisory Board Discussions)
- GitHub
 - Unit Repository code
 - QUDT code/files
 - Metrology ontology files



Challenges

- Workflow
 - What is the best workflow to administer unit representation creation and give appropriate authority/responsibility at all stages?
- Governance
 - How to ensure integrity of the process of unit representation creation?
 - How to ensure integrity of the repository?
- Engagement
 - How to get community involvement from appropriate stakeholders?
 - How to engage with the metrology community?