Harmonizing Units of Measure Vocabularies on the Web: an Initial Prototype

Kai Blumberg[1,2], Simon Cox[3], Hajo Rijgersberg[4], Chris Mungall[5], James Overton[6]

[1]Biological and Chemical Oceanography Data Management Office (BCO-DMO), Woods Hole Oceanographic Institution
[2]Department of Biosystems Engineering, University of Arizona, USA
[3]Molecular Ecosystems Biology, Lawrence Berkeley National Laboratory, USA
[4]Land and Water, Environmental Informatics, CSIRO, Australia
[5]Food Informatics and Supply Chain Management, Wageningen University & Research, Netherlands
[6]Knocean Inc, Canada

BCO-DMO
Biological & Chemical Oceanography Data Management Office

KNOCEAN

CSIRO

Image: https://www.atlasobscura.com/places/the-last-original-standard-metre-paris-france

FAIR

Findable
Accessible
Interoperable
Reusable

To humans

…

As well as machines

Applies to data and vocabularies

Wilkinson et al. 2016 *Scientific Data*

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles.

I3. (meta)data include qualified references to other (meta)data.

Existing Unit Vocabularies

UCUM Pattern for the unit symbols, widely used in biomedicine. In FHIR (Fast Healthcare Interoperability Resources).

https://ucum.org/

Notable mention: CODATA working group ‘Digital Representation of Units of Measure’

Gkoutos et al. 2012.	*FAIRsharing Team. 2015*	Rijgersberg et al. 2011	*Madin et al. 2007*	*Moncoiffe and Kokkinaki 2018*
UO	QUDT	OM	OBOE	NERC P06
Biology & Medical	**Engineering**	**Food Science & more**	**Ecology**	**Oceanography**
Developed by the OBO foundry for use with OBO ontologies e.g., PATO, GO.	Developed for the NASA Exploration Initiatives Ontology Models project, now 501c3.	Developed from several official standards, e.g., NIST. Model’s various scientific disciplines.	Developed at NCEAS to capture the semantics of scientific observation and measurement.	Extensive catalog of units used in Marine data. Part of NERC Marine Vocabulary system.

Example Generated Unit Term

rdfs:label [language: en]
metre per second

definition [language: en]
A unit which is equal to 1 metre per 1 second.

SI_code
m s−1

skos:exactMatch
http://qudt.org/vocab/unit/M-PER-SEC

skos:exactMatch
http://ecoinformatics.org/oboe/oboe.1.2/oboe-standards.owl#MeterPerSecond

skos:exactMatch
http://www.ontology-of-units-of-measure.org/resource/om-2/metrePerSecond-Time

skos:exactMatch
http://purl.obolibrary.org/obo/UO_0000094

skos:exactMatch
http://vocab.nerc.ac.uk/collection/P06/current/UVAA

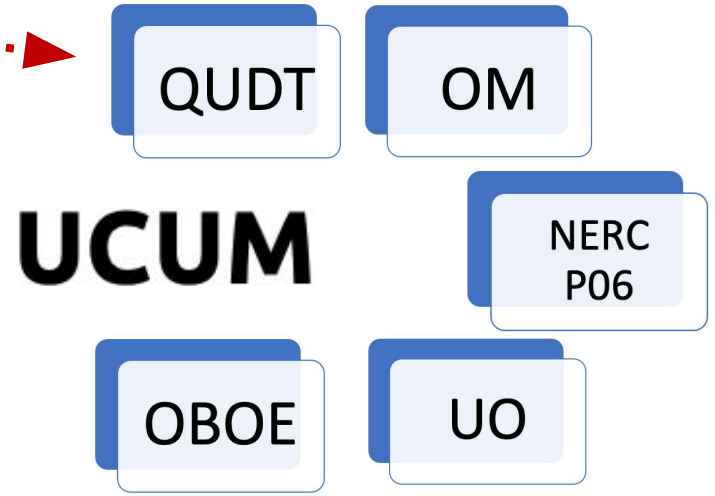
ucum_code
m.s−1

PURL will be: https://w3id.org/units/m.s-1

https://github.com/kaiiam/UO_revamp/tree/main/nc_name_script

Input: UCUM code
e.g., '**m.s-1**'
Output .ttl vocabulary

Maps to other systems from 1:1 UCUM code mappings



Protocol:

https://github.com/kaiiam/UO_revamp/tree/main/nc_name_script

Input UCUM code:

nmol.L-1.d-1

prefix: nano
base: mole

base: litre
exponent: -1

nmol **L-1** **d-1**

Capture constituent unit parts

Use base lookup tables to get info for definitions and labels

UCUM_symbol	label_en
A	ampere
cd	candela
K	kelvin
mol	mole
m	metre
s	second
g	gram

label_en	symbol
mega	M
kilo	k
hecto	h
deca	da
deci	d
centi	c
milli	m

Label: *"nanomole per day litre"*

Alphabetically sort to canonical UCUM code

nmol.d-1.L-1

Lookup UCUM code for mappings to unit systems

IRI	UCUM
http://qudt.org/vocab/unit/A	A
http://qudt.org/vocab/unit/YR	a
http://qudt.org/vocab/unit/B	B
http://qudt.org/vocab/unit/BARN	b
http://qudt.org/vocab/unit/BAR	bar

http://qudt.org/vocab/unit/NanoMOL-PER-L-DAY

Apply Standard URL escape character function to generate valid NC name for term ID

E.g.: nautical mile
[nmi_i]

%5Bnmi_i%5D

Generate final IRI e.g.: https://w3id.org/units/nmol.d-1.L-1

Statistics:

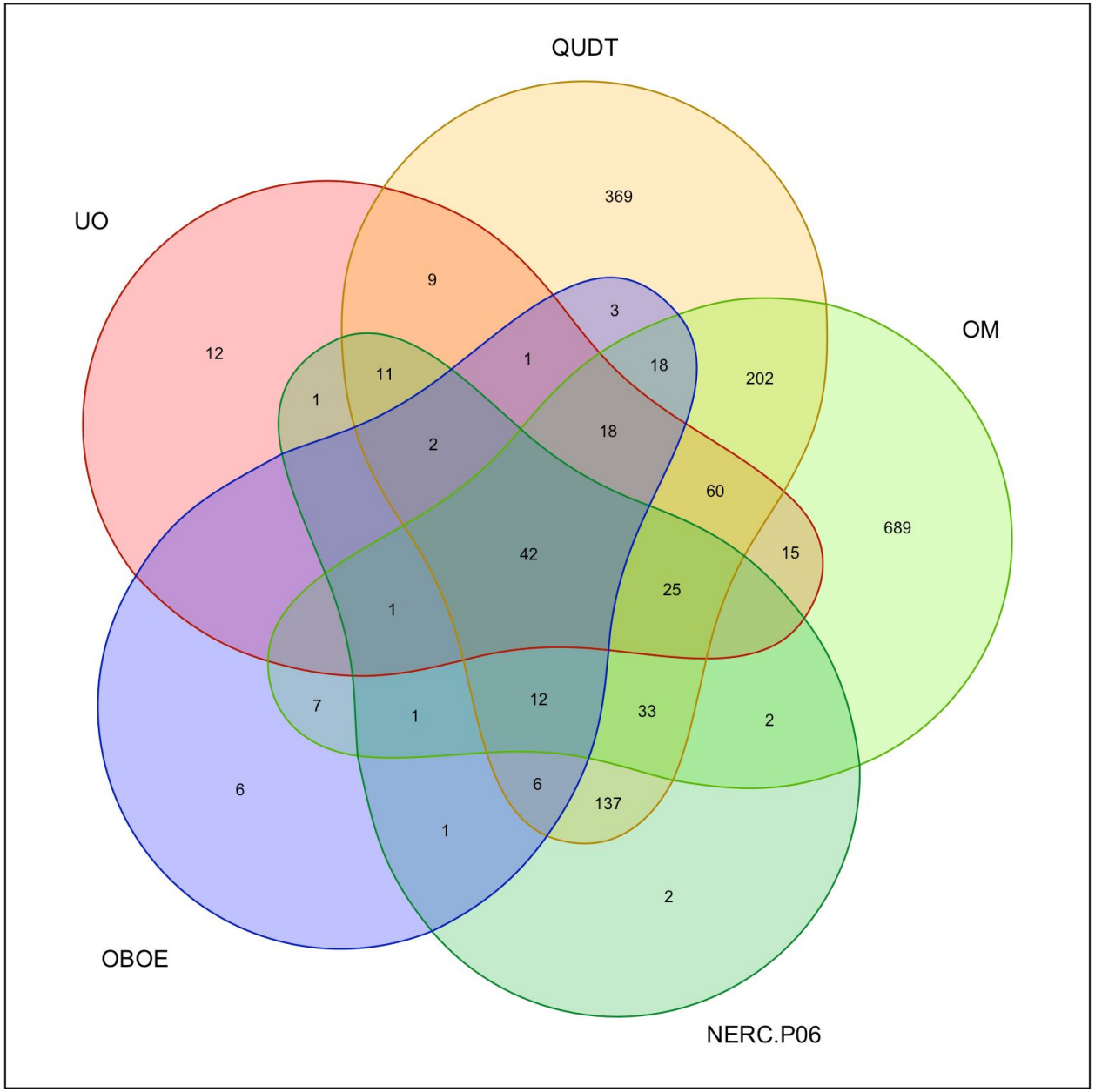
Number of unit terms currently in new system: **1947** (all unique)

	UO	QUDT	OM	NERC P06	OBOE
Number of unit terms: (not unique)	551	1685	2001	384	138
Total number of unit terms mapped to the new system: (above systems have duplicates e.g., nmol/L and nmolar)	235	1122	1221	383	133
Number of unique units mapped to new system: (not including duplicated concepts)	198	948	1126	277	119

Mapping Coverage

Showing the number of terms mapped to each existing system **within** the new system.

The new system currently has 1947 total units, 1685 of which are mapped to other system, thus 262 are unique to the new system (mostly metric prefix combos e.g., *“zettaneper”).*



Future Directions:

- Setup PURL server to run protocol:
  - Flask sever (or similar) for backend
  - Frontend UI will enable creation of new units based on UCUM codes
  - Dynamically run protocol to generate new units
    - Add new terms to ontology
- Add new features:
  - Additional languages Spanish, Chinese, French
  - Dimension vectors for dimensional analysis
- Finalize mappings from existing systems
  - E.g.: NERC, UO etc. to UCUM
  - Use mappings to populate initial term set

Questions?

UO | QUDT | OM | OBOE | NERC P06

rdfs:label [language: en]
metre per second

definition [language: en]
A unit which is equal to 1 metre per 1 second.

SI_code
m s-1

skos:exactMatch
http://qudt.org/vocab/unit/M-PER-SEC

skos:exactMatch
http://ecoinformatics.org/oboe/oboe.1.2/oboe-standards.owl#MeterPerSecond

skos:exactMatch
http://www.ontology-of-units-of-measure.org/resource/om-2/metrePerSecond-Time

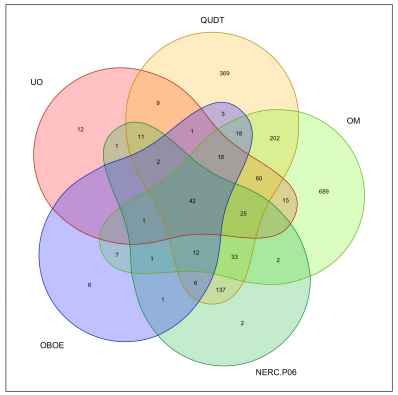
skos:exactMatch
http://purl.obolibrary.org/obo/UO_0000094

skos:exactMatch
http://vocab.nerc.ac.uk/collection/P06/current/UVAA

ucum_code
m.s-1

https://w3id.org/units/m.s-1

FAIR



the Unified Code for Units of Measure

prefix: nano
base: mole
base: litre
exponent: -1

nmol L-1 d-1

E.g.: nautical mile
[nmi_i]

%5Bnmi_i%5D

Supplemental Demo:

- Basic implementation of dynamic system:

https://units.ontodev.com/

- NOTE this is **NOT** a working system, for demonstration and alpha-testing only!