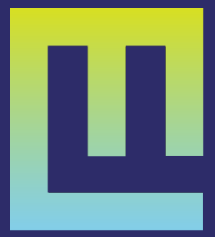


UNITS**ML**

Introducing UnitsML

CALCONNECT

UnitsML for semantic units

Our vision

- facilitate unit information processing
- validating consistent usage of units, quantities, and prefixes
- unit conversion and manipulation

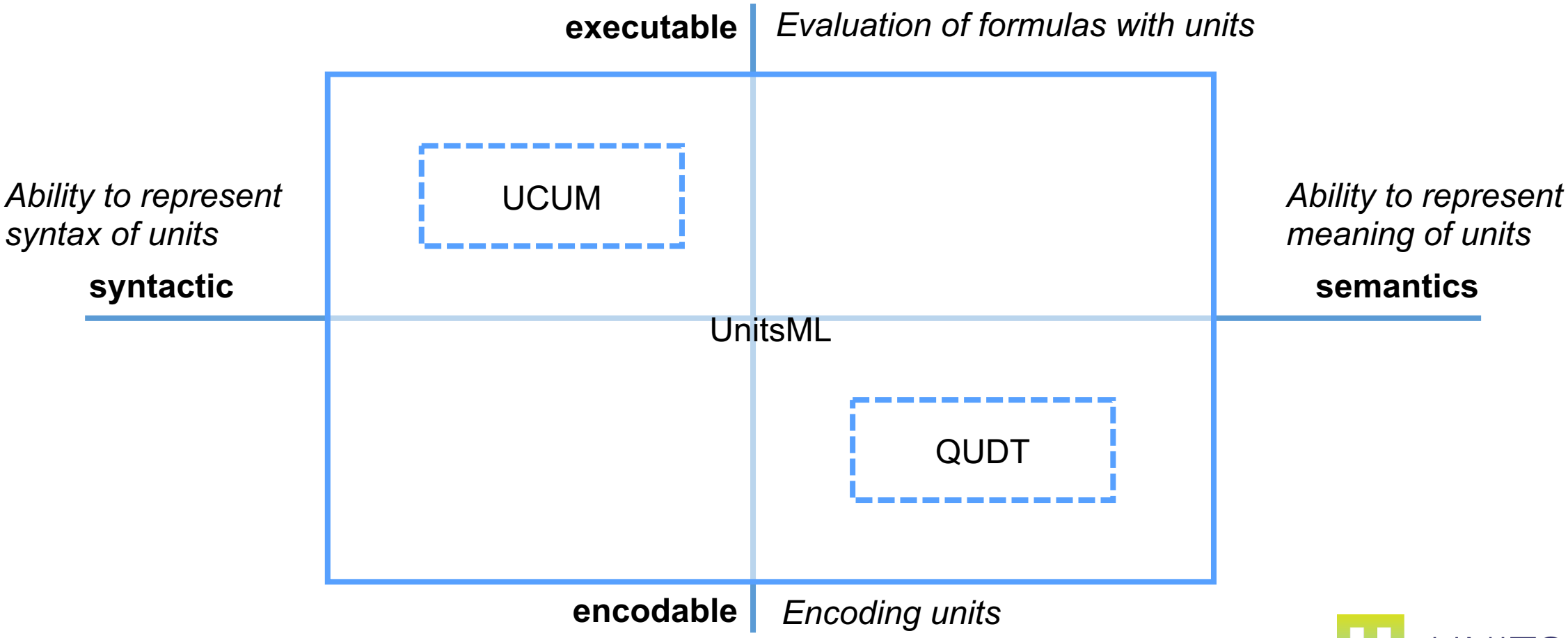
UNITSML

Digital units systems

- UCUM: representation of units and quantities in plain text
- QUDT: ontology of units, quantities
- UnitsML:
  - UnitsML XML: representation of units in XML
  - AsciiUnits: representation of units in plain text
  - UnitsML/MathML: representation of units with quantities in XML
  - AsciiUnits/AsciiMath: representation of units with quantities in plain text
  - UnitsDB: definitions, ontology of units

UNITSML

Digital units landscape



UNITSML

A history of UnitsML

- 1998: Initiated by Frank Olken, John McCarthy at LBNL
- 2003: UnitsML partially funded by NIST SIMA program, contributions from Bob Dragoset and many other NIST members
- 2006: UnitsML getting standardized at OASIS, proposed by NIST PML (previously PL) and sponsored by NIST
- 2016: OASIS TC UnitsML closed due to completion of work
- 2020: UnitsML work continues with NIST contributors, used by the BIPM SI Brochure's semantic version, new website
- 2022: CalConnect TC UNITS established to continue UnitsML work (joint TC with OASIS being discussed)

UNITSML

CalConnect TC UNITS

- TC UNITS is a standing committee chartered to provide international standards for:
  - representation and encoding of units, quantities, and dimensions;
  - mechanisms for encoding and interpretation of formulas that utilize units.
- The TC aims to work closely with the BIPM, CIPM and CGPM towards an interoperable “Digital SI” infrastructure, and advocate such throughout technologies and standards developed at CalConnect.
- Co-chairs: Karen Olsen (NIST), Nick Nicholas (Ribose)
- *In discussion with OASIS on a Joint TC (OASIS and CalConnect have a reciprocal membership agreement)*

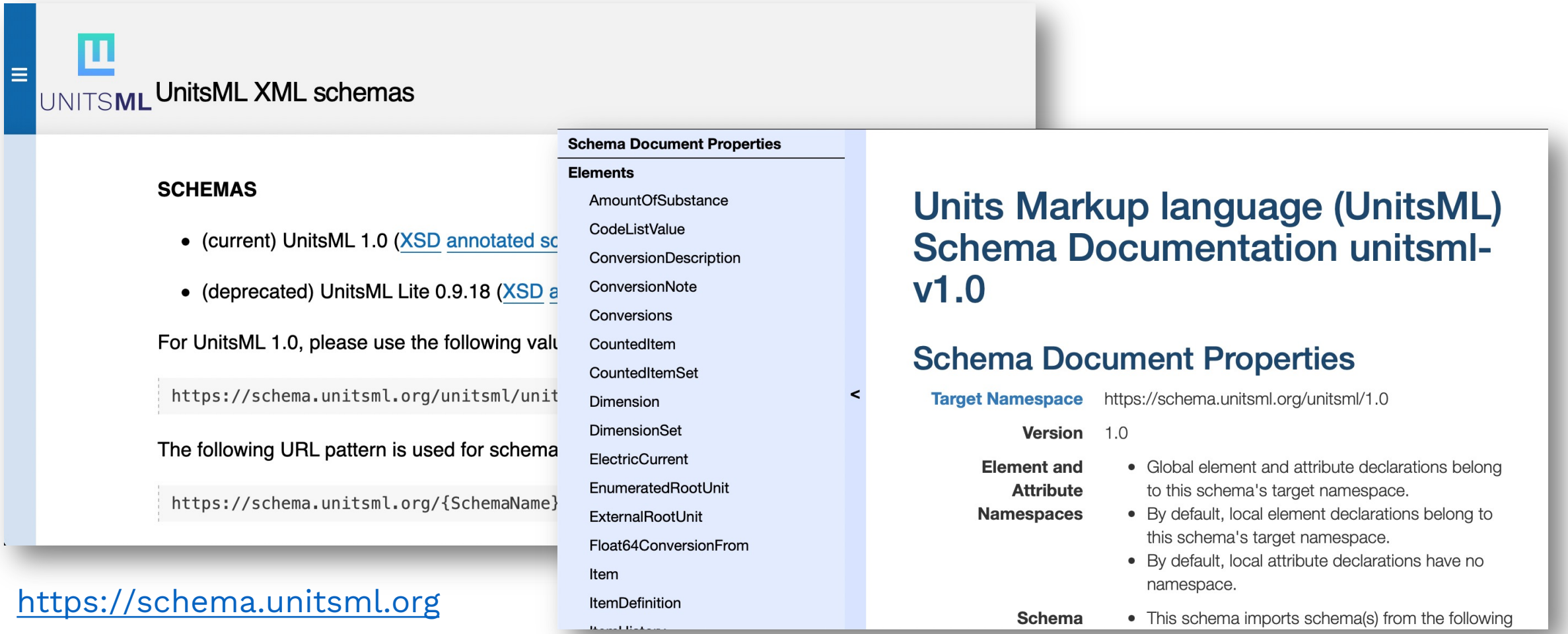
UNITSML

Components of UnitsML

- **UnitsML schema:** XML schema for encoding and constructing scientific units of measure
  - https://schema.unitsml.org
- **UnitsDB:** Data repository of known units of measure with encoding and rendering information on units, quantities, prefixes
  - https://github.com/unitsml/unitsdb
- **UnitsML tools:** Tools for parsing, rendering and manipulation of units in UnitsML

UNITSML

UnitsML XML schemas



https://schema.unitsml.org

https://schema.unitsml.org/docs/unitsml/unitsml-v1.0/

UNITSML

UnitsML works with MathML

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mn>0.012</mn>
  <mo rspace="thickmathspace">&#x2062;</mo>
  <mrow xref="U_kg.mol-1">
    <mi mathvariant="normal">kg</mi>
    <mo rspace="thickmathspace">&#x2062;</mo>
    <msup>
      <mrow>
        <mi mathvariant="normal">mol</mi>
      </mrow>
      <mrow>
        <mo>&#x2212;</mo>
        <mn>1</mn>
      </mrow>
    </msup>
  </mrow>
</math>
```


UNITSML

UnitsDB

- Managed dataset as ISO 19135 registry that provides information needed to reference units by machines
- Provides detailed units, dimensionality information for SI units and common non-SI units (>400):
  - unique identifiers
  - unit symbols
  - language-specific unit names
  - representations in terms of other units (including conversion factors)
- And unit systems, prefixes and quantities

UNITSML

Sources of UnitsDB

- BIPM SI Brochure
- NIST SP 811
- ISO 80000 series
- in some cases, "Encyclopaedia of Scientific Units, Weights and Measures, Their SI Equivalences and Origins" by François Cardarelli, is consulted

UnitsDB sample entry

```
"NISTu5":
  dimension_url: "#NISTd5"
  short: kelvin
  root: true
  unit_system:
    type: "SI_base"
    name: "SI"
  unit_name:
    - "kelvin"
  root_units:
    enumerated_root_units:
      - unit: "kelvin"
        power_denominator: 1
        power_numerator: 1
  quantity_reference:
    - name: "thermodynamic temperature"
      url: "#NISTq5"
```

```
unit_symbols:
    - id: "K"
      ascii: "K"
      html: "K"
      mathml: "<mi mathvariant='normal'>K</mi>"
      latex: \ensuremath{\mathrm{K}}
      unicode: "K"
    - id: "degK"
      ascii: "degK"
      html: "&#176;K"
      mathml: "<mi mathvariant='normal'>&#176;K</mi>"
      latex: \ensuremath{\mathrm{^{\circ}K}}
      unicode: "°K"
```


UNITSML

UnitsML tools

- UnitsML libraries for UnitsDB lookup and manipulation
- Converting AsciiMath

UNITSML

Also works with AsciiMath

Specified as:

```
0.012 "unitsml(kg*mol^(-1))"
```

Rendered as:

$$0.012\ \mathrm{kg\ mol^{-1}}$$

Encoded as:

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mn>0.012</mn>
  <mo rspace="thickmathspace">&#x2062;</mo>
  <mrow xref="U_kg.mol-1">
    <mi mathvariant="normal">kg</mi>
    <mo rspace="thickmathspace">&#x2062;</mo>
    <msup>
      <mrow>
        <mi mathvariant="normal">mol</mi>
      </mrow>
      <mrow>
        <mo>&#x2212;</mo>
        <mn>1</mn>
      </mrow>
    </msup>
  </mrow>
</math>
```

BIPM. The International System of Units (SI), Semantic Edition, 2022.

UNITSML

Real life example: SI Brochure

UNITSML

Bureau international des poids et mesures (BIPM)

- *the international organization established by the Metre Convention, through which Member States act together on matters related to measurement science and measurement standards*
- *the home of the International System of Units (SI) and the international reference time scale (UTC).*

- See more at https://www.bipm.org

UNITSML

SI Brochure: BIPM’s core deliverable

- Communicates the standards established by the Metre Convention, the international treaty that forms the basis for international agreement on units of measurement
- Basis of other international/national standards on usage of units (ISO and IEC 80000 series, NIST SP 330, etc.)
- Definitive terminology around units

UNITSML

ISO/IEC SMART levels indicate potential for functionality

A higher SMART level provides “more functionality”:

➔ more expressiveness allowed

➔ more accurate semantics

➔ more functionality possible

Level 1: Undifferentiated

undifferentiated digital content organized primarily for presentation and human readability without particular regard to semantics

EXAMPLE: A PDF file generated by a word processing system.

Level 2: Presentational

digital content organised according to presentational structure with some semantically organized content

EXAMPLE: An XML document that provides some structured terminology content.

Level 3: Presentational with semantics

digital content organised according to presentational structure with some semantically enriched content

EXAMPLE: An XML document that contains structured terminology content and linked references to external terminology content.

Level 4: Semantically oriented

digital content organised according to a semantics-oriented structure

EXAMPLE: An ontological representation of requirements.

Level 5: Native semantics

digital content organised according to their natural or native semantic structure

EXAMPLE: Information models in EXPRESS and a registry of such data.

UNITSML

The Semantic SI Brochure uses UnitsML

- BIPM developed a machine-readable semantic version of its SI Brochure as a SMART Level 4 output
- Used Metanorma, an open-source model-based authoring suite for standards
- Available open source at [warning: DRAFT!]:
  - https://github.com/metanorma/bipm-si-brochure

UNITSML

Level 1: Undifferentiated digital content

Table 3. Base quantities and dimensions used in the SI

Base quantity	Typical symbol for quantity	Symbol for dimension
time	t	T
length	l, x, r, etc.	L
mass	m	M
electric current	I, i	I
thermodynamic temperature	T	Θ
amount of substance	n	N
luminous intensity	I_v	J

BIPM. The International System of Units (SI), Edition 9, 2019.

- Some math symbols do not exist in the font needed.
- Some math symbols were represented by images.
- Reusability: very difficult.

UNITSML

Level 2: Presentational

Table 6. Examples of SI coherent derived units whose names and symbols include SI coherent derived units with special names and symbols

Derived quantity	Name of coherent derived unit	Symbol	Derived unit expressed in terms of base units
dynamic viscosity	pascal second	Pa s	$kg\ m^{-1}\ s^{-1}$
moment of force	newton metre	N m	$kg\ m^{2}\ s^{-2}$
surface tension	newton per metre	$N\ m^{-1}$	$kg\ s^{-2}$
angular velocity, angular frequency	radian per second	$rad\ s^{-1}$	s^{-1}
angular acceleration	radian per second squared	$rad\ s^{-2}$	s^{-2}
heat flux density, irradiance	watt per square metre	$W\ m^{-2}$	$kg\ s^{-3}$

BIPM. The International System of Units (SI), Edition 9, 2019.

- Some math symbols are represented by normal, italicized or bolded text.
- The typical way to insert math symbols.
- Reusability: less difficult but at the cost of reimplementing semantics.

UNITSML

Level 3: Presentational with semantics

Presentation MathML

```
<mi mathvariant="normal">J</mi>
<mo rspace="thickmathspace">&#x2062;</mo>
<mi mathvariant="normal">s</mi>
```

Rendered in Level 2:

- **the Planck constant,** h**, is** $6.626\,070\,15 \times 10^{-34}$ J s,

- Presentation MathML
- Reusability: copyable from XML, the meaning of the unit is lost.

BIPM. The International System of Units (SI), Semantic Edition, 2022.

UNITSML

Level 4: Semantics oriented

Presentation MathML with UnitsML

```
<mrow xref="U_NISTu13.u3">
  <mi mathvariant="normal">J</mi>
  <mo rspace="thickmathspace">&#x2062;</mo>
  <mi mathvariant="normal">s</mi>
</mrow>
```

Rendered in Level 2:

- **the Planck constant,** h**, is** $6.626\,070\,15 \times 10^{-34}$ J s,

Reusability: copyable from XML, semantics fully preserved.

UnitsDB entry

```
"NISTu13":
  dimension_url: "#NISTd15"
  unit_system:
    type: "SI_derived_special"
  unit_name:
    - "joule"
  unit_symbols:
    - id: "J"
  root_units:
    enumerated_root_units:
      - unit: "joule"
        power_denominator: 1
        power_numerator: 1
  quantity_reference:
    - name: "Lagrange function"
      url: "#NISTq152"
  si_derived_bases:
    - id: NISTu1
```

BIPM. The International System of Units (SI), Semantic Edition, 2022.

UNITSML

Level 5: Native semantics

Content MathML with UnitsML

```
<csymbol cd="unitsml">U_NISTu13.u3</csymbol>
```

- Reduction of a math formula as digital data.
- Reusability: copyable from XML, semantics fully preserved.

Rendered in Level 2:

- **the Planck constant,** h**, is** $6.626\,070\,15 \times 10^{-34}$ **J s,**

BIPM. The International System of Units (SI), Semantic Edition, 2022.

UNITSML

Join us to make UnitsML better!

Current members:

- Mark Carlisle
- Simon Frechette
- Peter Linstrom
- Nick Nicholas
- Karen Olsen
- Ronald Tse
- Mike Douglass

UNITSML

Thank you!

UNITSML

Comparing with UCUM

- Ability to represent formulas with units and quantities in more than one textual format (ASCII, XML and potentially JSON)
- Supports the kilogram as a basic unit (UCUM uses “gram”)
- Allows users to create “custom units” built on SI units (or other UnitsDB units)
- UnitsML suite covers everything that UCUM can do, “UnitsML with AsciiMath” is an ASCII-based math format like UCUM. It could support any representation format in the future, e.g. LaTeX.

UNITSML

Comparing with QUDT

- QUDT is an ontology, it is not meant to facilitate the representation of formulas or execution of them

UNITSML