LTER Unit Dictionary: Best Practices

Version 14 – 05/27/2011

Introduction

With a growing understanding of the central role of units in defining scientific data, the Long-Term Ecological Research (LTER) Information Management Committee (IMC) is addressing a challenge associated with integrating environmental data by developing a Unit Dictionary including Best Practices on formulation and use of scientific units within the community. When units are well defined and created following common practices, it is possible to convert and evaluate interrelations of measurements.

In a distributed scientific community, the definition of a unit is ultimately up to an individual or a group of participants since a unit may be proposed, defined and used locally at any time. The intent of LTER Unit Dictionary project is both to preserve this autonomy by permitting any unit to be used as a site-specific unit and to facilitate standardizing activities at the same time. A LTER Unit Dictionary contributes to both local and larger-scale unit standards-making efforts.

Once a community dictionary is in place, a need arises for an accepted interpretation of use of the dictionary in particular situations. Identifying and articulating that common interpretation furthers communication among community participants about accepted practices and helps to ensure the dictionary reflects experiences with unit creation and use. Aiming to make unit use explicit and to support unit creation amongst the LTER sites, this document is offered as a guide to LTER conventions and definitions.

Unit Dictionary Best Practices Summary

Unit Assignment Best Practices

1. Encourage use of authoritative units whenever possible.
2. Ensure that assigning a unit is appropriate for the data being described.
3. Review existing unit standards for names and abbreviations, considering existing community units before creating a new unit.
4. Place only broad and essential measurement information with the unit metadata.

Unit Name Best Practices

1. Identify first the unit then the modifier.
2. Singular terms are preferred over plural terms.
3. Use “Per” as a linking term in the unit name.
4. Use camelCase notation for readability.
5. Do not include numbers or special characters in the unit name.
6. Non-metric coefficients are acceptable.

Abbreviation Best Practices

1. Use characters in the UTF-8 character set.
2. Use a standard notation for exponents and ratios.
3. Use multiplication symbols for clarity.

Description Best Practices

1. Use the description field to provide information for using and interpreting the unit.

Parent Best Practices

1. SI parent units are preferred over non-SI parent units.
2. Supply multiplier and constant for converting from the child unit to the parent unit.
3. Use a multiplier of 1 and a constant of 0 to show equivalence.

Management Best Practices

1. Units that are incorrect or unused should be deprecated, not deleted.
2. Quantities (unit types) are currently considered optional information.

Discussion

Each of the Best Practices is documented below using description, examples, and references where available. Further background is provided by an annotated resource list, a National Institute of Standards and Technology checklist, and a related document titled "LTER Unit Dictionary: History and Context."

Unit Assignment Best Practices

1) Encourage use of authoritative units whenever possible.

Granting that the choice of units is a scientific decision and outside of the immediate Information Management realm, it is best to encourage use of authority units if there is a unit choice open for IM input. 'Authority' units here are approved SI or ISO-31 units for instance. Using accepted units supports data standardization and integration.

2) Ensure that assigning a unit is appropriate for the data being described.

Classifying your data into measurement types can be helpful in assessing the need to designate a unit. A five-category classification system outlines the following measurement types: 1. interval (any measurement without a meaningful zero, i.e. degrees Celsius), 2. ratio (any measurement that a calculation can be performed on), 3. datetime (times with a variety of formats), 4. nominal (text fields i.e. text station name), and 5. ordinal (values with a code-like significance, i.e. pH). Of these five, only two measurement types - interval and ratio - call for specification of units.

A common issue arises when an attribute is given an incorrect measurement type that references a unit; i.e. an attribute is described as measurementType "interval" or "ratio" which require units, but should be datetime, nominal or ordinal, which are not associated with units.

Incorrect: A sampling date is entered with a unit such as "yyyy", or "month".

Correct: Dates (and date parts) should be given measurement type "datetime" with the pattern "yyyy" indicated elsewhere in the documentation; no unit is used with dates.

Reference: See IMC EML Best Practices

3) Review existing unit standards for names and abbreviations, considering existing community units before creating a new unit.

When searching for an existing unit or thinking about creating a new unit, it is best to check whether a standard unit is available in the LTER unit dictionary or in an outside standard. Existing site metadata records may not include a standard unit due to many factors, including legacy units developing at sites independently without easy reference to existing units, or a unit created and assigned before a new dictionary version is released.

Incorrect: The unit "meter" is in the dictionary but a custom unit "m" (for meter) was created.

Correct: Creation of a new unit should be reserved for units that do not currently exist in dictionaries, or units with a local meaning or use different from dictionary entries.

References: LTER Unit Dictionary: History and Context (Standards section)

4) Place only broad and essential measurement information with the unit metadata.

Units are associated with attributes, which are commonly referred to as variables, parameters or columns in the environmental sciences. However, there remains a gray area in making a distinction as to where the boundary lies between unit and attribute information. Further, there is lack of consensus in terms of attribute names and how to approach describing an attribute more fully. In local scientific conventions and practices, there is often a blending or overlap between units and attributes. On a structural level and for an unambiguous comparison of measurements, the attribute and unit must be distinguished.

Incorrect: The candidate was called a custom unit, but (to the best of our knowledge) is really an attribute (Example 1, below), or part of the attribute was included in the unit name (Examples 2 and 3). The correct unit may have been already available as a standard unit (e.g. 'number' for Example 1).

Correct: All attribute information must be removed from the unit and stored elsewhere.

Example 1:

Attribute: bacterial abundance

Correct Unit: number (num)

Improper unit: bacterial abundance

Example 2:

Attribute: primary production

Correct Unit: milligramPerMeterCubedPerDay (mg/m3/day)

Attribute-level information: carbon

Improper unit: milligramCarbonPerMeterCubedPerDay (mgC/m3/day)

Example 3:

Attribute: short shoot growth

Correct Unit: numberPerCentimeterSquared (num/cm2)

Improper unit: short shoots per cm2

Reference: See Rule NIST#11 at http://www.physics.nist.gov/cuu/Units/checklist.html

Unit Naming Best Practices

1) Identify first the unit then the modifier.

Modifiers such as 'squared' and 'cubed' are common in ecological units. For correct unit naming, the unit is listed first with any modifiers following. This applies to each element in the unit name if multiple base units are brought together into a derived unit.

Incorrect: gramPerSquareMeterPerSquareSecond

Correct: gramPerMeterSquaredPerSecondSquared

Reference: no authoritative standard identified

2) Singular terms are preferred over plural terms.

Though in speech we can easily adapt our terminology to singular and plural situations, standardized unit notation is not situational. All units should be created using singular terminology rather than pluralized.

Incorrect: grams, metersSquared

Correct: gram, meterSquared

Reference: see NIST SP 1020-1

3) Use "Per" as a linking term in the unit name.

Linking terms are necessary parts of a complex unit name as they express the relationship between the base units brought together. Using only 'Per' standardizes the unit relationships allowed while minimizing characters used in the unit name.

Incorrect: gramDividedByMeterSquared, meter over hour

Correct: gramPerMeterSquared

Reference: This convention is commonly used in SI, NIST and other organization unit websites and documents; the authoritative source is unknown at this time.

4) Use camelCase notation for readability.

By removing spaces while preserving some level of human-readability through the capitalization of the first letter in terms, camelCase notation creates machine- and code-friendly terms that minimize the number of characters and are recognizable across platforms. Capitalization is used for readability only, not for string comparison. If a unit is modified with a prefix (kilo, centi, etc), the prefix and unit are treated as one term rather than two (i.e. the unit is not capitalized as a second term after a prefix). Modifiers (Squared, Cubed, etc) however, are considered separate terms for camelCase consideration.

Incorrect: count per meter squared, kiloGram

Correct: countPerCentimeterSquared

Reference: no authoritative standard identified for camelCase notation; prefix+unit combination as a term is based on SI case of 'kilogram'

5) Do not include numbers or special characters in the unit name.

Unit names should be fully spelled out with no abbreviations or special characters used. This maximizes the ability of a machine to automatically recognize and compare units through their names.

Incorrect: metersPerSecond2

Correct: meterPerSecondSquared

Reference: no authoritative standard identified

6) Non-metric coefficients are acceptable.

Non-metric coefficients (OneTenth, FourHundreth, etc.) are acceptable in unit names. Units with coefficients should not be pluralized. All coefficients should be written out and not abbreviated, in keeping with practice 5 above.

Incorrect: numberPer1000MeterCubed

name should not contain numbers

Incorrect: numberPerThousandMetersCubed

unit should not be pluralized

Correct: numberPerThousandMeterCubed

Reference: no authoritative standard identified

Description Best Practices

1) Use the description field to provide information for using and interpreting the unit.

The unit description provides a field where information regarding the use and interpretation of the unit can be included. This field can be used to summarize information in other unit fields in a more human-readable format. The information required will vary from unit to unit, so there is no set format for a proper unit description. Some information that is often useful includes:

- A description of the quantity (unit type) represented by this unit
- A reference to the source of the unit, if applicable
- Information on conversions between this unit and other unit, including but not limited to this unit's parent unit
- An explanation of any non-metric coefficient in the unit name
- Examples of usage of the unit in specific domains

Example: **picoCuriesPerGram:** pCi/g = 1E-12 Curies per gram of sample, equal to 2.22 radioactive disintegrations per minute per gram of sample

Example: **gramPerFourHundrethMeterSquare:** grams of substance per four hundredths of a square meter, 20 cm x 20 cm quadrat

Example: **microeinsteinPerMeterSquaredPerSecond:** micro Einsteins (1E-06 moles of photons) per square meter per second (radiant flux density), irradiance

Reference: no authoritative standard identified

Abbreviation Best Practices

1) Use characters in the UTF-8 character set.

Abbreviations, unlike unit names, may contain numbers and special characters. For compatibility reasons, unit abbreviations should be limited to characters in the UTF-8 character set. UTF-8 is fully compatible with ASCII, so any ASCII encoded strings are valid UTF-8 strings as well.

2) Use a standard notation for exponents and ratios.

When representing exponents in a unit abbreviation, one of two notations should be used:

- Superscript notation (e.g. m^2)
- Caret notation (e.g. m^2)

Ratios should be indicated using one of two notations as well:

- Divisor notation (e.g. m/s)
- Negative exponent notation (e.g. ms^{-1} or ms^-1)

Abbreviations should not mix the two types of exponent notation or the two types of ratio notation.

Incorrect: meterPerSecondSquared abbreviated as m/s2

Exponents should use superscript or caret notation

Correct: meterPerSecondSquared abbreviated as m/s^2 or m/s^2

Incorrect: micromolePerMeterSquaredPerSecond abbreviated as mol • m^{-2}/s

Different types of notation should not be mixed

Correct: micromolePerMeterSquaredPerSecond abbreviated as mol/m^2/s or mol • m^{-2} • s^{-1}

Reference: no authoritative standard identified

3) Use multiplication symbols for clarity.

Abbreviations for complex units can be difficult to read. To improve readability, it is acceptable to include a • symbol to indicate multiplication of two elements. This also helps to distinguish elements that are multiplied from multi-character abbreviations such as kg, mol, or Pa.

Example: newtonMeterSecond abbreviated as N• m • s rather than nms.

Reference: no authoritative standard identified

Parent Best Practices

1) SI parent units are preferred over non-SI parent units.

When choosing a parent unit, an SI unit should be chosen if possible. An SI unit is defined as a unit that is the product of the seven SI base units (meter, kilogram, second, ampere, Kelvin, mole, and candela) with integer exponents. Standardizing on SI units as parents makes conversions between units of the same type simpler.

Incorrect: nanogramsPerHourSquared with a parent of gramsPerHourSquared

Neither gram nor hour are SI base units, so gramsPerHourSquared is not an SI unit

Correct: nanogramsPerHourSquared with a parent of kilogramsPerSecondSquared

Kilogram and second are the SI base unit for mass and time respectively, so kilogramPerSecondSquared (also called a newtonPerMeter) is the SI unit for surface tension

Reference: no authoritative standard identified

2) Supply multiplier and constant for converting from the child unit to the parent unit.

A multiplier and a constant should be given such that, when applied to a measurement in this unit will convert it to a measurement in the parent unit. The multiplier should be applied first, followed by the constant. If no multiplier is needed, the multiplier should be set to 1. If no constant is needed, the constant should be set to 0.

Incorrect: millimeter with a parent of meter and a multiplier of 1000

*1 millimeter * 1000 = 1000 meters is not a correct conversion*

Correct: millimeter with a parent of meter and a multiplier of .001

*1 millimeter * .001 = .001 meter is a correct conversion*

Reference: STMML schema definition

3) Use a multiplier of 1 and a constant of 0 to show equivalence.

Sometimes a unit of measurement will have multiple names by which it can be used. To show that a unit is equivalent to an SI unit, but with a different name, set the SI unit as the alternately named unit's parent, and set the multiplier to 1 and the constant to 0.

Example: **newton** is equivalent to kilogramMeterPerSecondSquared

Example: **joule** and **newtonMeter** are equivalent to kilogramMeterSquaredPerSecondSquare

Reference: no authoritative standard identified

Management Best Practices

1) Units that are incorrect or unused should be deprecated, not deleted.

Units should not be deleted from the LTER unit dictionary, even if they are found to be incorrect or unused. Instead, these units should be flagged as deprecated, with a preferred unit indicated. This preserves the history of the changes made to the LTER unit dictionary, and also makes it easy out of date units to be found and resolved.

2) Quantities (unit types) are currently considered optional information.

Quantity (called unit type in the STMML schema) is a field that represents the general category of phenomena measured by a unit. For example, meterPerSecond, inchPerYear, and milePerHour are all measurements of the 'speed' quantity. Quantity information is useful fro determining the relationship between units, and whether conversions from one unit to another are possible. Quantity also implies some restrictions – for example, a unit's parent should always be of the same quantity. However, the definitions and usage of various quantities are not well documented, so as of this revision of the best practices, quantity information is considered optional and will not be explicitly covered by the best practices.

Annotated Resource List

A NIST reference on constants, units and uncertainty provides a bibliographic reference, a guide to SI units, and a document on unit conventions.

http://www.physics.nist.gov/cuu/index.html

http://physics.nist.gov/cuu/Units/bibliography.html

http://www.physics.nist.gov/cuu/Units/checklist.html

The Bureau International des Poids et measures (BIPM) website with a SI document with guidelines for creating units and also two non-SI units lists giving recommended (i.e. time, plane, angle, area, volume, and mass) and non recommended units.

http://www.bipm.org/en/si/si_brochure/

An article introducing STMML a markup language for scientific, technical and medical publishing.

http://www.jstage.jst.go.jp/article/dsj/1/0/1_128/_article

A list of standardUnits included in EML2.0.1.

http://cvs.ecoinformatics.org/cvs/cvsweb.cgi/eml/eml-unitDictionary.xml

CUAHSI units and other controlled vocabularies:

http://water.usu.edu/cuahsi/odm/cv.aspx

as well as the CUAHSI Units and Observation Data Model (ODM):

and best practices for adding data to CUAHSI is provided:

http://his.cuahsi.org; http://river.sdsc.edu/Wiki/Default.aspx