OpenMath Content Dictionaries for SI Quantities and Units

Joseph B. Collins

Naval Research Laboratory
4555 Overlook Ave, SW
Washington, DC 20375-5337
joseph.collins@nrl.navy.mil

Abstract. We document the creation of a new set of OpenMath content dictionaries to support the expression of quantities and units under the International System of Units (SI). While preserving many of the concepts embodied in the original content dictionaries, these new content dictionaries provide a foundation for quantities and units that is compliant with international standards. We respond to questions raised in prior efforts to create content dictionaries for units and dimensions by proposing and applying some rationalized criteria for the creation of content dictionaries in general. The results have been released and submitted to the OpenMath website as contributed content dictionaries.

1 Introduction

We are interested in the creation of a scientific markup for representing physics based models. In pursuing this objective, we have not found a sufficient body of markup for creating documents representing physics based models, rather we find we must develop further markup constructs in order to create such documents. Luckily, we find that we can build upon a developing body of work in mathematical markup. The first step in this endeavor, from a bottom-up perspective, is to properly address the representation of quantities and units.

OpenMath [1] represents a significant effort amongst the various attempts at representing mathematical knowledge, particularly in the problem area of representing mathematical semantics using web-oriented standards. Scientific knowledge is a mixture of mathematical representations and references to experiments and measurements. One of the fundamental intersections of this dual nature of scientific knowledge is in the representation of quantities and units. An initial attempt to capture some units and physical dimensions [2], [3] has resulted in several OpenMath content dictionaries (CDs). The prior OpenMath CDs we refer to are: dimensions1, units_metric1, units_imperial1, units_us1, units_time1, units_siprefix, units_ops1, and physical_consts1. In these efforts, some attention has been paid to observing conventions specified in the International System of Units, or Le Système International d'Unités, hereinafter simply referred to as SI, as expressed in [4], [5], and [6]. In these efforts there is an admitted incompleteness with respect to adherence to the SI in the implementation of the associated OpenMath CDs for units and dimensions.

Our work here is an attempt to improve the representation of dimensions and units within OpenMath by building an new set of explicit SI CDs, following the current SI standard [7]. We do this in an attempt to close the gap between OpenMath CDs for units and dimensions and the expression of the formal standards of units and dimensions embodied in the SI. In this paper, we document a proposed restructuring of OpenMath CDs for units and dimensions, as well as attempt to provide answers to questions raised in prior efforts.

1.1 Guiding Principles

Prior to beginning this exposition, we review some of the guiding principles that we have developed and used. First, the SI represents a few things: a long-standing, slowly evolving international consensus; a well developed set of coherent standards for quantities and units; and a standard that dominates other, similar standards. Other standards, such as Imperial units and United States units, many of which predate the SI, have, over time, typically been redefined in terms of SI standards, and are largely in a process of being officially phased out. One of the purposes of the SI is to provide, from a few defined standard units, a way to repeatably measure all physical quantities. As such, the SI is, foremost, a standard based on physical measurement, not a mathematical standard. The definition of quantities, units, and their properties is essentially a posteriori, following from observation. While it is not generally anticipated that there would be major changes in these observations, in principle, if there were a distinct change, the meanings of what are defined in the SI would change. This being said, there is an observed mathematical structure to the concepts of quantities and units. This observed mathematical structure constitutes a physical theory, rather than mathematical truth: our acceptance of its truth rests primarily on consistency with observation and measurement. Being mathematical in nature, a physical theory's truth requires mathematical consistency, but the simplicity or complexity of its mathematical nature does not determine its validity. Neither can the mathematical nature of a physical theory be determined a priori. Much of the mathematical structure of quantities and units is described in the SI, and we intend to capture it as far as possible within the OpenMath framework.

While most non-specialists are familiar with the concept of units, the concept of quantities is perhaps somewhat more esoteric, particularly as the SI addresses it. Most importantly, some care is required in discussions of quantities and units, as colloquial usage may often be incorrect with respect to the SI. Many preconceptions may have to be abandoned for old habits to be replaced with SI compliant usage.

While embarking on this effort we were faced with several issues regarding how to make best use of OpenMath CDs. OpenMath and Content MathML [8] both embrace the concept that it is not only possible, but desirable, to separate the expression of mathematical semantics from the expression of the *presentation* of various mathematical symbols commonly in use. We advocate observing this distinction and avoiding the temptation to mix them more than might be necessary. In the development of content markup, we also suggest applying the following three criteria, as defined below: lack of ambiguity, convenience, and simplicity.

Ambiguity is anathema to content markup. Different from presentation markup, content markup loses utility if there is ambiguity in semantics. We suggest that ambiguity in mathematical semantics is tolerable, even desirable, in presentation markup. Presentation options of a given concept may overlap with presentation options of other concepts, potentially giving rise to ambiguity. In presentation, ambiguity is usually balanced with economy in the space allotted to symbols, relying on a context of implicit conventions. Presumably, a coherent set of presentation options can, and should, be documented in a style guide and perhaps implemented using a style sheet based translator operating on a document having content markup. Such an accompanying style guide can alleviate problems of ambiguity in the presentation markup.

By simplicity in content markup, one thing we mean is that we try to avoid implementing redundant constructs. While having multiple ways to express the semantics of a particular concept is not wrong, simplicity in a content markup language, i.e., having only one principal way to express a certain concept, seems desirable. Simplicity means fewer symbols to remember and will otherwise aid in the use of content markup, for example, in minimizing the job of writing translators for presentation of content. In Content MathML and OpenMath, operators and functions are currently represented in prefix form without much redundancy. For example, a prefix *divide* symbol represents the concept of division. While a presentation markup certainly does support alternative symbolic expressions of the operation of division, it is unnecessary for content markup to do so. Infix variations of division operators are not needed in content markup. Consequently, we do not insist that content markup include the various symbolic representations of division: an obelus symbol, a vinculum symbol, or a virgule or solidus symbol. By contrast, we suggest that this type of simplicity is not as important in presentation markup as it is in content markup. Presentation options of particular concepts need to be as numerous as the presentation conventions one intends to support, such as all of the above representations of the operation of division.

This does not mean there should never be any overlapping semantics: *divide* is redundant with exponentiation with negative integers. We find this redundancy acceptable because it is convenient to support both *divide* and *power* symbols, and it can be done without ambiguity. By convenience we mean that we may implement a possibly redundant set of concepts, usually because it is easier to do so than not to do so. Clearly, there are trade-offs between simplicity and convenience and the decision as to what is right and proper is subjective: the overarching consideration is overall economy of effort for the whole enterprise of implementation and use of the markup.

In sum, when faced with a question of what is the best way to represent a particular concept, we find we must first consider whether it relates to: a scientific or measurement issue; a presentation issue; or a mathematical semantic issue. Secondly, in deciding whether to represent a particular concept, we must consider how to justify it based on principles of simplicity, convenience, and resolving ambiguity. We apply these guiding principles in the following sections.

2 SI Quantities

Quantities are a fundamental concept within the SI, where quantity is defined [7] as a "property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed by means of a number and a reference". This *reference* means a reference amount of the same kind of quantity, called a unit. The magnitude is the number of unit amounts of the quantity required to be equal to the amount possessed by the object of interest.

The concept of quantities arises in multiple physical contexts, from the very specific to the abstract. For example, we can refer to the length of the specific piece of furniture standing in front of us, the wavelength of an arbitrary frequency of light, or the general concept of length. The objective is to be able to express all of these. From the SI definition, one can see that the concept of quantity is complex, and must have a combination of properties, which include a *dimension*, a *kind*, a *unit*, and a *magnitude*. The dimension, kind, and unit properties each have label, or name, values, and the dimension and unit properties also have SI symbol values, essentially abbreviated aliases of the names. A magnitude must have a numerical value expressed in some unit. These are discussed below.

2.1 Quantity Dimension

The most primitive concept of physics markup is that of *physical dimension*. With this property, mathematics is transformed to enable representation of physical quantities. We here define physical dimension, according to common usage, to include that which is defined as the *quantity dimension* in the SI standard. In this usage, the term physical dimension, such as is used in dimensional analysis, refers not only to the SI quantity dimension, but also to general SI quantities, including SI derived quantities.

The use of the term *dimension* in SI, as it relates to physical dimension, is, however, much more restrictive. The quantity dimensions in systems of quantities and units are given by the products of powers of a set of *base quantities*. As such, the base quantities form a basis for the space of quantity dimensions. Any system of quantities can, using its own set of base quantities, define a basis with which to span some space of physical dimensions. For the SI, there are seven base quantities: length, mass, time, electric current, thermodynamic temperature, amount of substance, and luminous intensity. The full set of SI quantity dimensions are generated by products of powers of these specific base quantities. Physical dimensions that are not in this set of SI quantity dimensions are not referred to as dimensions within SI. A different system of quantities might define a separate set of base quantities, consequently having a different set of quantity dimensions.

Derived quantities, as defined by SI, are quantities, in a system of quantities, which are defined in terms of the base quantities of that system. The SI requires that all quantities in the system be defined in terms of a product of powers of the base quantities. The SI introduces the concept of a mapping, dim, which maps a derived quantity to a quantity dimension. The requirement that all SI quantities be defined in terms of products of powers of base quantities essentially constitutes the definition of this mapping, dim. The SI does not bound the number of derived quantities that may be introduced.

Mathematical Structure of Quantity Dimensions The mathematical structure of quantity dimensions in SI is summarized as follows: There are seven *base quantities*, length, mass, time, electric current, thermodynamic temperature, amount of substance, and luminous intensity. The *base quantities*, with the inclusion of an eighth, variously named *quantity of dimension one*, or *dimensionless quantity*, form an abelian generating set for the infinite abelian group of objects variously referred to as *quantity dimensions*, *dimensions of a quantity*, or just *dimensions*. The group multiplication operator is compatible with the multiplication operator for the field of real numbers. In addition to the base quantities, there is an unbounded set of *derived quantities*, which, in the SI, are defined by a name and a non-injective mapping to a quantity dimension,

$$\mathsf{dim} : \mathit{derived\ quantities} \rightarrow \mathit{quantity\ dimensions}. \tag{1}$$

When two quantities, $Q1$ and $Q2$, are said to be *dimensionally equivalent*, this means that $\mathsf{dim}(Q1) = \mathsf{dim}(Q2)$. In addition to multiplication, each quantity dimension may be raised to any real power, though only rational powers ever seem necessary.

2.2 Kind of Quantity

The next property of a quantity is that of the *kind* of quantity. The kind of a quantity distinguishes between different quantities that may have the same quantity dimensions. The SI concedes that the concept of the kind of a quantity is to some extent arbitrary. Nevertheless, it is a necessary distinction. Perhaps the best illustration is by way of examples. The salinity of a solution is typically stated as a mass fraction, i.e. the mass of dissolved salt per unit of mass of solution. As such, salinity is a dimensionless quantity, i.e., $\mathsf{mass}/\mathsf{mass} = 1$. Angle is also dimensionless, given, for example, in radians as the ratio of the length of the subtended circular arc and the radius of the same circle. While dimensionally equivalent, one still considers salinity and angle to be distinct kinds of quantities. There are many dimensionless quantities distinguished by kind. Similarly, the quantities *torque* and *energy* have the same quantity dimension but are distinguished from each other by being different kinds of quantities.

Similar to the mapping, dim, we introduce the concept of another mapping, kind, implicit in the SI, which maps a quantity to a quantity kind. Complete equivalence of two quantities, $Q1$ and $Q2$, can only occur when both

$\mathsf{dim}(Q1) = \mathsf{dim}(Q2)$ and $\mathsf{kind}(Q1) = \mathsf{kind}(Q2)$ are true. While the SI recognizes that quantities may differ in kind, it does not standardize the definition of kind.

Distinction of quantities by kind can have rather fundamental consequences. The distinction between *inertial mass*, the resistance to acceleration due to an applied force in Newton's second law, and *gravitational mass*, the proportional factor by which an object influences other objects through the force of gravity in Newton's law of universal gravitation, was once debated. This debate was famously settled with Einstein's principle of equivalence and subsequent experimental measurements. We discuss later the import of differences in kinds of quantities.

2.3 Units

A *unit* is a defined reference amount of a given quantity. Having a unit, any quantity of the same kind may then be expressed as equivalent to some numerical amount of the reference quantity. For example, an arbitrary mass may be expressed as a numerical amount times a reference amount of mass. In the SI system of units, there are seven base units: the metre, kilogram, second, ampere, kelvin, mole, and candela, corresponding to the seven SI base quantities. There is a necessary one-to-one mapping between the base quantities and the base units. An additional unit, one, is added, corresponding to the dimensionless base quantity.

The essential reason for identifying the base units is that they serve as the measurement standards for most physical measurements. If, for example, we wanted to measure the length of something, we would need to calibrate our length-measuring device using a standard length. In SI, that standard length is the metre, and its definition is in terms of a measurement procedure. The same is true for all of the base units: each is defined in terms such that they may be readily used to calibrate measurement equipment for the corresponding physical quantity. This is referred to as the *practical realization* of the base units.

Similar to the dim mapping, the SI also posits a unit mapping with a range that is the set of products of powers of the SI base units, i.e., the set of *coherent derived units*. The domain of the unit mapping is the set of all quantities defined within the SI. As there is for each derived quantity a mapping to an SI quantity dimension, so too is there a mapping to a coherent derived unit. Because of this, the meaning of a quantity is defined in physically measurable terms, i.e., the base unit definitions. There can be no circular definitions, or any definitions of measurable quantities that are not rooted in definable measurements.

A limited number of coherent derived units are given special names. These named coherent derived units are: radian, steradian, hertz, newton, pascal, joule, watt, coulomb, volt, farad, ohm, siemens, weber, tesla, henry, degree Celsius, lumen, lux, becquerel, gray, sievert, and katal.

Mathematical Structure of Units: The mathematical structure of units in SI is summarized as follows: There are seven *base units*, metre, kilogram, second, ampere, kelvin, mole, and candela. The *base units*, with the inclusion of the unit named *one*, with symbol, 1, form an abelian generating set for the infinite abelian group of objects called the *coherent derived units*. The multiplication operator of this group of coherent derived units is compatible with the multiplication operator for the field of real numbers.

There is a one-to-one mapping of quantity dimensions to coherent derived units,

$$\mathsf{unit} : \mathit{quantity\ dimensions} \rightarrow \mathit{coherent\ derived\ units}. \quad (2)$$

The unit of any quantity, Q, is given by $\mathsf{unit}(\mathsf{dim}(Q))$.

2.4 Magnitude of a Quantity

The magnitude of a specific scalar physical quantity represents the amount of that quantity, i.e., a mathematical product of a real number and a reference quantity, or unit. For example, the kilogram is the reference mass that resides at the International Bureau of Weights and Measures (BIPM) in Sèvres, France. In the SI system, the masses of all other physical objects are measured in proportion to that standard kilogram, where the proportion is expressed as a limited precision real number. There is also an accompanying error value, representing an estimate of the standard deviation, were an ensemble of such measurements to have been conducted. By default, when not specified, the error in a number is assumed to be half of the place value of the least significant digit expressed.

The SI introduces the concept of an operator that returns the numerical value of a quantity. We denote this mapping, num:

$$\mathsf{num} : \mathit{quantities} \rightarrow \mathit{real\ numbers}. \quad (3)$$

Any quantity, Q, may be represented in the system by the unique product $\mathsf{num}(Q) \cdot \mathsf{unit}(Q)$. Clearly, while an arbitrary quantity is independent of unit system, num and unit are specific to a particular system of units, e.g., the SI.

Vector quantities, including complex numbers and tensors, are used in some situations. A full treatment of these is not possible here, partly because OpenMath currently expresses Cartesian vectors, but not vectors in other coordinates, e.g., spherical coordinates. When vector quantities are used, they may either be heterogeneous, where each component may be a distinct quantity, or homogeneous, where the components are pure numbers and the unit may be represented as a distinct factor. It should be noted that multiplication of base units with vector magnitudes is very much like scalar multiplication: when vector quantities fail to commute, it is the num parts of the quantities that fails to commute, not the unit parts. Extending the definition of the num mapping is straightforward for the much more common case of homogeneous Cartesian vectors of quantities, for example,

$$\mathsf{num}((3\mathsf{m}, 5\mathsf{m}, 7\mathsf{m})) = (3, 5, 7). \quad (4)$$

The representation of standard error with vector quantities is typically in terms of a covariance structure, a subject not elaborated in the SI, or here.

3 Operations on Quantities and Units

There are several operations that are defined on quantities and units. There are those that are specific to a system of quantities and units, namely dim, kind, unit, and num. The SI explicitly defines the dim, unit, and num mappings. In the SI, the unit mapping is symbolically represented with square brackets, and the num symbol is represented with curly braces. We add to this the kind mapping which provides some distinction between quantities which may be dimensionally equivalent. The equivalence of two quantities requires the equivalence of all four of these properties, i.e., two arbitrary quantities, $Q1$ and $Q2$, are equivalent if and only if $\mathsf{dim}(Q1) = \mathsf{dim}(Q2)$, $\mathsf{kind}(Q1) = \mathsf{kind}(Q2)$, and $\mathsf{num}(Q1) = \mathsf{num}(Q2)$. (If $\mathsf{dim}(Q1) = \mathsf{dim}(Q2)$, then $\mathsf{unit}(Q1) = \mathsf{unit}(Q2)$). The SI also specifies the multiplication, division, and raising of quantities to powers as required for the meaningful construction of quantities and units. If a quantity, say Qx, is expressed in a non-SI unit, say Ux, conversion may be effected by having defined $\mathsf{unit}(Ux)$ when Ux itself is defined. Such conversions must be linear transformations: affine, or additive conversion, such as the conversion of thermodynamic temperature from kelvin to degree Fahrenheit, are not defined within this space of units and these operators. On the other hand, conversion of kelvin to degree Rankine is well defined.

3.1 Other Operations

There are several natural language usages which, with respect to dimensions and units, imply various mathematical operations. We partition these into the following categories: one, times, per, and plus; SI prefixes; cube, cubic, square, and squared.

One, Times, Per, and Plus: The dimensionless base quantity, one, and its corresponding unit, one, are essentially synonymous with the mathematical symbol of the same name. The presence of the unit or quantity dimension one in a quantity expression that we may want to represent in markup is generally implied, though rarely required explicitly. In text, multiplication is implied by juxtaposition of quantity names, using either a space (invisible times), a dot symbol, a hyphen, etc., and encoding such expressions into markup should follow those implications. As we have stated, multiplication of base quantities and base units is associative, commutative and otherwise compatible with multiplication of the real numbers, so the order of the encoded terms with respect to each other and numerical scalars is, in general, not semantically significant.

The term *per*, as in metre per second, implies an infix division operator, and could be encoded as such as long as it is unambiguously used.

The arbitrary addition of quantities, as pointed out for temperature, for example, in [3], is not always physically meaningful. In general, the SI does not discuss addition of quantities. We point out here that addition of quantities is undefined for quantities that are not of the same kind, without leaving out the possibility of additional constraints on addition of quantities. Similarly, the arithmetic relational operators equals, greater than, less than, etc., are only meaningful when quantities are of the same kind. The question as to whether quantities are of the same kind is really only answered by experimental validation of a law of physics. A law of physics is typically phrased as an arithmetic relation of quantities, implying the additivity of quantities involved. Arbitrary relations of quantities may be mathematically well defined without possessing physical meaning. Sometimes, as in the case of temperature, it is possible to identify different kinds of temperature, e.g., absolute (thermodynamic) temperature and relative temperature. The SI allows the definition, presumably by any user, of such derived quantities. The assignment of a value to the kind mapping in the definition of any such quantity, is also, presumably, left to the user, as is the task of ensuring the validity of the physical semantics.

SI Prefixes: SI prefixes are normally used by prepending a single prefix to an SI base unit, with the exception of the kilogram. For mass quantities, the prefix is prepended to the term *gram* as if it were the base unit, though gram is not presented without such a prefix. SI prefixes act like multiplicative numerical constants, each an integer power of 10. Normally, it is the user's choice to express, for example, 1000 microgram, 1 milligram, or 10^{-6} kilogram. Whichever choice is of no consequence for semantic content.

For capture of the semantic value of SI prefixes, we suggest the following: a) the use of an empty element for each occurrence of a named prefix; b) the allowance in content markup of multiple occurrences of prefixes (forbidden by SI in presentation); c) the abandonment of an OpenMath *prefix* symbol in the units_ops1 CD in preference to the multiplication operator, *times*. The enforcement of the SI-required, normal form of only using single prefixes is a presentation constraint, without mathematical meaning. It may be recommended practice in content markup, but should not be required.

Cube, Cubic, Square, Squared The derived units *square metre*, *cubic metre*, *metre squared*, *metre cubed* are cases where the terms *square* and *cubic* act like prefix exponent operators, and *squared* and *cubed* act like postfix exponent operators. If used, a hyphen is interpreted as an infix multiplication operator. Similarly, *metre per second* and *metre/second* are cases where the symbols *per* and "/" act like infix division operators. In each of these cases, unless there is a compelling reason for supporting infix, postfix, or alternative, redundant prefix operators, simplicity and maintaining compatibility with the existing OpenMath mathematical CDs with *times*, *divide*, and *power* symbols suggests using the existing prefix forms of these same operations.

3.2 Appending of Units

While it may, in the expression of a quantity in markup, be good practice to append units to a numerical expression, this has no mathematical semantic value.

The associativity and commutativity of multiplication of units with each other and with the numerical magnitudes makes the meaning clear regardless of order. The SI requirement to express the units of a quantity after the numerical value is strictly a prescription of a standard presentation form. Its practice may be recommended, but should not be required in content markup.

4 Number Representations

There are three requirements for numerical representations of quantities: range of absolute magnitudes, precision, and uncertainty. Magnitudes of physical constants and their uncertainties range from 10^{-72} to 10^{50} in SI units. The known material universe is estimated to be made up of approximately 10^{80} nucleons. Expressed in Planck units (not typically used for everyday physics) the overall size of the known universe is on the order of 10^{62} Planck lengths, the age is on the order of 10^{62} Planck times, and the mass of the order 10^{62} Planck masses. Any number representation should be able to support expression of numbers of these magnitudes. The current upper limit of precision of measurement for physical constants is about twelve to thirteen significant decimal digits, such as for the Rydberg constant.

While it may be ideal to have a format for arbitrary precision and arbitrary magnitude measurable values, IEEE double precision format [9] provides about 16 significant decimal digits and magnitudes spanning 10^{308} to 10^{-308}. While it is not inconceivable that physics-based computations may exceed the dynamic range the IEEE double precision format, for most practical purposes, that format appears adequate for the present.

Unambiguous representation of a limited precision real number requires the use of a scientific notation with significand of limited length and an exponent. Scientific notation is needed to express, for example, exactly two significant digits of 6.2×10^3, instead of writing 6200, the latter being ambiguous. Depending on how the significand of a double precision format is interpreted, the semantics of the precision limit of a measurement can be lost in the conversion to and from machine double precision. Machine formats do not typically support an inherent precision of a number which is distinct from the machine precision. From a semantic perspective, the machine precision is merely the maximum precision that may represented in a machine word, not the actual precision of the value that needs to be represented.

The IEEE double precision format does not in itself support the expression of measurement precision, which is really a form of uncertainty. Considering that in the conversion of a double precision literal to a machine double the limit of precision expressed in the literal is usually lost, it appears necessary to provide a separate mechanism for explicitly expressing the limit of precision, or for expressing the uncertainty. As such, we find it necessary to represent the concept of standard error for representing measured quantities, i.e., an estimate of the standard deviation, as a way to capture adequately the normal expression of uncertainty for scalar numerical magnitudes. Interpretation of uncertainties as bounded ranges, for example, is not standard practice. Proper treatment of uncertainty is sufficiently complex that we do not here provide a solution to this requirement, but suggest that more information, either an accompanying integer to represent the number of significant digits, or an accompanying float magnitude, will be needed to represent the standard error.

5 Physical Constants and Other Measured Quantities

Historically, the standardization of units began with units of practical interest, and as such are inherently anthropocentric. Over time, these have been supplanted with units having sizes that are still anthropocentric, but being defined in ways more amenable to increasingly precise scientific measurement. Nevertheless, any measurement device will have bounded precision. As a result, there are many scientifically measurable quantities that are difficult to capture adequately using only SI units.

The SI base units provides a set of units for representing the bulk of everyday quantities. It is possible, however, due to limits of precision in measurement apparatus, that macroscopic standards, such as the kilogram, are inadequate for measuring very small or very large quantities, such as the masses of quantities smaller than a microgram. It is known, for example, that the standard kilogram varies on the scale of micrograms, so it is not well defined to a corresponding precision. Electron and nucleon masses are significantly smaller than this. To adequately represent measurements of quantities having a combination magnitude and precision that fall outside the SI base units, and measurement equipment that uses them, the SI provides some other, off-system, measured units. The values of these in terms of SI units are obtained experimentally. These units are the electronvolt, the dalton, and the astronomical unit as units of energy, mass, and length, respectively.

An alternative system of units that is frequently used by physicists, for example, in studying cosmology and quantum gravity, are Planck units. The magnitudes of Planck units and the measurement apparatus required to gauge them make them inappropriate for use in anthropocentric applied physics. Planck units are sometimes called "God's units", as they comprise a natural or intergalactic standard set of units, completely defined using universal constants. The five Planck units are the Planck length, Planck mass, Planck time, Planck charge, and Planck temperature. They are completely defined in terms of the following five universal physical constants: the Newton's gravitational constant, the reduced Planck constant, the speed of light, the Coulomb force constant, and the Boltzmann constant. They are largely defined by experimental measurement, and updated measured values are published periodically by, for example, the international organization CODATA. These Planck units and the constants that define them are intrinsic to the physical laws that appear to describe the origins of the universe and physics as we know it. Consequently, other unit systems and fundamental constants are seen as derivative with respect to these.

6 Proposed Modifications to the OpenMath CD library

We agree with prior work on the definition of quantity dimensions and units as empty XML elements. This appears to be their most natural representation. We do, however, recommend the reorganization of the dimension and unit CDs appropriate to the properties of the SI quantities and units.

We believe that SI quantity and unit CDs should be specifically labeled as such, and the base units and quantities separately identified. Other systems having their own base quantities and units should similarly be distinctly labeled. Non-SI units and quantities should not be freely intermixed with SI quantities and units, but should be coupled in some way so as to distinguish them from units defined with respect to a different system. For example, while the contents of the dimensions1 CD may all be thought of as dimensions, in SI there are only seven base quantities serving as *quantity dimensions*. Other quantity dimensions are products of powers of these base quantities, i.e., derived via mathematical rules of construction, and generally need not be individually defined. For these reasons, we introduce the SI_BaseQuantities1 CD and suggest the deprecation of any SI base quantities from the dimensions1 CD. The SI_BaseQuantities1 CD defines length, mass, time, current, temperature, amount of substance, and luminous intensity. We additionally introduce the corresponding SI_BaseUnits1 CD and suggest the deprecation of SI base units from the existing CDs where they occur. The SI_BaseUnits1 CD defines metre, kilogram, second, ampere, kelvin, mole, and candela. For completeness, in both the SI_BaseQuantities1 and SI_BaseUnits1 CDs we include the symbol *one*, equating it to the symbol of the same name found in the OpenMath alg1 CD.

As the SI base quantities and units are defined as standards, the definition and a reference to the defining documents should be explicitly cited within those CDs. These definitions are generally sufficiently brief to be described within the OpenMath Description elements. Typically, new unit standards are ratified by a standards body at a periodically held conference, so a citation should, for example, name the conference and year. A change in SI definitions should prompt new versions of these CDs.

Derived quantities are those that are defined in terms of the base quantities. Some derived quantities are defined in the SI, particularly when they have names that are not mathematically constructed, or have corresponding specially named units. While the SI admits an unlimited number of derived quantities, and can only specify rules for their creation, it is reasonable to identify as SI quantities and units those that are specifically mentioned in the standard. For this reason we introduce SI_DerivedQuantities1, which defines angle, solid angle, frequency, force, pressure, energy, power, charge, voltage, capacitance, resistance, conductance, magnetic flux, magnetic flux density, inductance, Celsius temperature, luminous flux, illuminance, radioactivity, absorbed dose, equivalent dose and catalytic activity which all have named SI units.

Correspondingly, we define in SI_NamedDerivedUnits1 the radian, steradian, hertz, newton, pascal, joule, watt, coulomb, volt, farad, ohm, siemens, weber, tesla, henry, degree Celsius, lumen, lux, becquerel, gray, sievert, and katal. We also include in SI_NamedDerivedUnits1 the exceptional implied unit, gram. We include the gram because, even though its SI compliant presentation requires use of a prefix, the gram is the semantic root of all SI mass units. In the interest of simplicity, we believe that unit names that reflect mathematical construction, such as metre_per_second, or metre_squared, should not be included in CDs.

SI_DerivedQuantities1 may, in principle, include all derived quantities that are compliant with SI. There is no mathematical bound in the number of such quantities that may be defined. OpenMath custom is to not make content dictionaries arbitrarily large. No clear means of restricting the size in general suggests itself. We include in SI_DerivedQuantities1, somewhat arbitrarily, the quantities area, volume, speed, momentum, moment of force, density, concentration, heat, and entropy along with those quantities that have their own SI units, simply because these additional quantities are specifically mentioned within the SI.

6.1 SI Symbols

Each SI base quantity and base unit, and the named units corresponding to some derived quantities, have names with prescribed spellings, as well as associated SI symbols which act as abbreviated names, such as "m" for metre. In the spirit of first identifying what is necessary for semantic capture and in the spirit of simplicity, we use the SI names of quantities and units for their OpenMath symbol names in the CDs, choosing those over the short, SI symbols as being less ambiguous and less subject to errors in usage. In OpenMath, the presentation of many symbols is not emphasized. For example, the gradient operator does not reference the nabla symbol. Accordingly, it seems that we can, in general, neglect the representation of the semantically redundant SI symbols in OpenMath CDs. As a compromise, we suggest that symbols be identified in the *Description* element of the *CDDefinition* for reference purposes.

6.2 Capitalization and Abbreviation

Persons names are not capitalized in SI unit names, with the single exception of "degree Celsius". Person's names are capitalized when used in quantity names. Otherwise capitals are only used when the dimension and unit names begin a sentence. The SI symbols for units may have capitals: one should consult the standard for specific values. Abbreviations of unit and quantity names are explicitly barred. While these are principally presentation considerations, we follow these conventions in the symbol definitions within the CDs.

6.3 Non-SI Units

There are many non-SI units in use, with varying degrees of status with respect to the SI. There are the following four categories of units: coherent SI units; non-SI units accepted for use with the SI; non-SI units that have been deprecated; other non-SI units. Off-system, or non-SI units are those that are not coherent SI units. For example, minute, hour, and day are off-system units, defined in terms of the second of time and retained for use with the SI. The degree, minute, second, and gon of arc, and the litre are all defined in terms of SI units and retained for use with SI units. Similarly, the measured units mentioned earlier, the electronvolt, the dalton, and the astronomical unit, are off-system units, but used with SI. We propose two CDs to represent the category of non-SI units accepted for use with the SI: SIUsed_OffSystemUnits1 and SIUsed_OffSystemMeasuredUnits1. We choose to distinguish the Planck units and the physical constants that define them in their own CD, called UniversalPhysicalConstants.

We decline to propose CDs for other non-SI units at this time. Instead, we suggest some criteria for their construction. Many, if not most, non-SI units have been redefined in terms of SI units. Units not defined or mentioned in the SI should appear in non-SI CDs with their corresponding definition in terms of SI units. For example, the most common *foot* is the international foot. Since the international foot overrides other foot standards, we would recommend its OpenMath symbol be merely named *foot*, not *international foot*. In 1958 the United States and countries of the Commonwealth of Nations defined the length of the international foot to be equal to 0.3048 metre. This should be defined using formal mathematical property statements, and both the unit and num symbols, in this case, $\mathsf{unit}(\mathsf{foot}) = \mathsf{metre}$ and $\mathsf{num}(\mathsf{foot}) = \mathsf{0.3048}$. (Following this convention of defining non-SI units will allow straightforward unit conversions). Other, less common, foot units may have an OpenMath symbol distinguished in name, such as the *United States survey foot*. In any case, an OpenMath Description element should make it unambiguous which unit is being identified by citing an appropriate standards document, as described above. Other units that are both non-SI and undefined in terms of SI units must, of metrological necessity, be defined within other unit systems. Other unit systems may be rooted in their own content dictionaries. Units defined within other unit systems in general will have no exact, mathematical conversions to SI units, only approximate, metrological conversions.

6.4 Interaction with Other OpenMath Content Dictionaries

OpenMath symbols applicable to quantities and units from existing OpenMath CDs include: zero, one, divide, minus, plus, power, product, root, sum, times, unary_minus, eq, eqs, lt, gt, neq, leq, geq, and approx.

In [2] an interesting proposal was made for using the names of physical dimensions as types for units. Due to questions, we do not here provide the Small Type System (STS) CDs for the SI units and dimensions. Certainly the SI dimensions would appear to qualify as being of type PhysicalDimension. One could possibly create a new, more restrictive sub-type of PhysicalDimension, called SI_Dimension. As for an STS type for units, if we were to follow the suggestion of using SI dimensions as types for units, there would be an infinite number of possible types, i.e., certainly all of the possible quantity dimensions. Certainly such a type system would be well structured. We have the means to construct any of them, but not to list them all. It is not clear why a single type, say *Unit*, would not suffice. One of the purposes for having each physical dimension, or quantity dimension, be a type for units, is to perform type checking. This purpose, as well as unit checking and unit conversions for different units of the same quantity dimension, would seem to be adequately served with the use of the mappings, kind, dim, unit, and num.

7 Summary

In total, we propose eight new OpenMath CDs, which have been released and submitted to the OpenMath website as contributed content dictionaries. Our proposed SI CDs for quantities, SI_BaseQuantities1 and SI_DerivedQuantities1, and the proposed criteria for accepting dimension definitions, largely make obsolete the prior dimensions1 OpenMath CD. Our proposed SI CDs for units, SI_BaseUnits1, and SI_NamedDerivedUnits1, and the proposed criteria for accepting unit definitions, largely make obsolete the prior units_metric1 OpenMath CD. We do not suggest replacement of the units_siprefix CD, though we suggest deprecation of the units_ops1 CD. We incorporate the above described dim, kind, unit, and num symbols into a CD named SI_Functions1. Our proposed FundamentalPhysicalConstants1 is redundant with some symbols defined in physical_consts1, but also introduces the non-SI Planck units. SIUsed_OffSystemUnits1 replaces minute, hour, day of time in units_time1. SIUsed_OffSystemMeasuredUnits1, does not affect existing CDs.

We do not comment on any particulars regarding the units_imperial1 and units_us1 CDs, at this time, other than to say that since they have not been explicitly defined in terms of SI unit CDs, we may want to redefine them in CDs, with appropriate citations, in terms of SI units, if appropriate standards organizations have done so.

Overall, these new CDs, as proposed, isolate the essentially physics-based SI content into two CDs: the SI_BaseUnits1, and the SIUsed_OffSystemMeasuredUnits1. The additional FundamentalPhysicalConstants1 is also essentially physics-based, as would be any definition of measured constants or non-SI base units of other systems. The other CDs are principally mathematical in nature, where any of their properties attributable to the measurement process is derived by association to symbols defined in the physics-based CDs.

Finally, we have, for lack of space and time, neglected to elaborate on the general representation of vector quantities and the uncertainties of quantities.

References

1. S. Buswell, O. Caprotti, D.P. Carlisle, M.C. Dewar, M. Gäetano, and M. Kohlhase. The OpenMath Standard 2.0. http://www.openmath.org, 2004
2. Davenport, J.H., Naylor, W.A.: "Units and Dimensions in OpenMath", (2003), http://www.openmath.org/documents/Units.pdf

3. Stratford, J., Davenport, J.H.: "Unit Knowledge Management", Intelligent Computer Mathematics, Lecture Notes in Computer Science, pp. 382-397, Springer Berlin / Heidelberg (2008) ISBN 978-3-540-85109-7
4. Organisation Intergouvernementale de la Convention du Mètre. The International System of Units (SI), 8th ed., http://www.bipm.org/utils/common/pdf/si_brochure_8_en.pdf
5. IEEE/ASTM. SI 10-1997 Standard for the Use of the International System of Units (SI): The Modern Metric System. IEEE Inc (1997)
6. Taylor, B.N.: Guide for the Use of the International System of Units (SI) (May 2007), http://physics.nist.gov/Pubs/SP811
7. Technical Committee ISO/TC 12, "Quantities and units", DRAFT INTERNATIONAL STANDARD ISO/DIS 80000-1, International Organization for Standardization, 2008
8. Content MathML Version 2: http://www.w3.org/TR/MathML2/chapter4.html, Content MathML Version 3 (Draft): http://www.w3.org/TR/MathML3/chapter4.html
9. IEEE 754-1985, see for example, http://en.wikipedia.org/wiki/IEEE_754-1985