UnitDim: an ontology of physical units and quantities

Hajo Rijgersberg[1], Jan Top[123]

[1] Agrotechnology & Food Innovations, Wageningen UR, P.O. Box 17,
NL-6700 AA Wageningen, The Netherlands
{Hajo.Rijgersberg, Jan.Top}@WUR.nl
http://www.ato.dlo.nl

[2] Vrije Universiteit Amsterdam, Faculty of Science, Computer Science,
De Boelelaan 1081A, NL-1081 HV Amsterdam, The Netherlands
http://www.few.vu.nl

[3] Wageningen Centre for Food Sciences, P.O. Box 557,
NL-6700 AN Wageningen, The Netherlands
http://www.wcfs.nl

Abstract. Many people are convinced that a formal, digital definition of units and quantities is required. Automated support for parameter and unit selection, dimension and unit consistency checking, conversion of units, et cetera, would benefit from such a standard and eliminate many errors and misconceptions in science and engineering. Already for many decennia, more or less formalized versions of different unit standards exist on paper. However, their translation into formal, digital representations is still inadequate. First, in our opinion, most proposals are inadequate towards the conceptual and technical aspects of units, quantities and dimensions. Second, the most serious efforts are not available in modern and widely accepted semantic languages such as RDF or OWL. Third, most approaches lack comprehensiveness. In this paper, we present a new ontological approach to the organization of units and related concepts. This new ontology, named UnitDim, is based on the existing paper standards, in particular, but not restricted to, the SI. UnitDim claims to model unit and quantity matters in an adequate way with respect to a number of practical services in supporting scientific and engineering activities. The ontology, written in OWL DL, presently contains more than 200 quantities and over 300 units. Besides describing the basic structure of the ontology, we also set out some intricate modeling issues and design decisions in this paper.

1 Introduction

The scientific community has always been a driving force for innovation in communication technologies, the (Semantic) Web being an outstanding example. However, only now the reverse effect is getting proper attention in what is called *e-science*. Due to a number of developments, e-science will influence scientific and engineering practice profoundly in the near future. Firstly, because scientists are moving from free text documents to digitized, structured information that can be processed by automated systems. Mathematical models are already a common way to express scientific knowledge. In fact, most emphasis in automating scientific practice has been on numerical processing and visualization. However, mathematics *sec* is not rich enough to support meaningful communication. The next challenge is to formalize the physical concepts and their relations underlying these models to enable proper interpretation. The mathematical expression $E = mc^2$ has no meaning if the domain context (what is the meaning of the variables, under which conditions is the expression relevant, what kind of objects do the quantities refer to, which units can be used?) is not given. Some semantics is encoded in modern software tools for model construction and analysis, but only limited effort has been spent on independent standards for semantically enriched scientific information sharing.

The second reason why e-science is getting attention nowadays is that the interaction between scientists has become much more intensive, crossing disciplinary boundaries, already at an early stage of research. Rather than exchanging finalized work through publications, world wide electronic communication allows exchanging early model proposals or raw data. This will significantly influence the dynamics of scientific research. It also implies the need for standards for exchanging (the semantics of) models and data.

So, we have observed a need to formalize scientific knowledge. A natural starting point for doing so is the *physical unit*. Many units have already been semi-formally defined on paper, in particular within the International System of Units (SI). A number of attempts have been made to construct digital formalizations [1-5]. However, at closer inspection it appears that these descriptions are either incomplete, not presented in terms of standard ontology languages or only presented in terms of typical examples. It is quite surprising how intricate a seemingly simple framework of physical units can be. We submit that a proper approach requires 1) profound knowledge of the fundaments of physical unit systems as defined by the major standardization bodies, 2) knowledge and skills to apply modern ontology languages and 3) practical experience with the application of units in research. The latter condition implies that an ontology of units serves a number of practical goals in science and engineering. In this paper we start from these practical issues to construct a fairly comprehensible unit ontology in terms of OWL DL [6].

It is important to note that we consider UnitDim as a first step towards formal semantics in e-science. The next step will be to show how quantities and their mutual relations and values can be formalized. Simultaneously to the development of UnitDim, we are developing ModelDat, an ontology that aims to offer a format for formulating and storing *statements*, usually referred to as models and data sets. UnitDim is supposed to be imported in ModelDat.

In the next section we provide an overview of the main concepts in UnitDim. In Sec. 3 the notion of *physical quantity* is described, being a necessary requirement for introducing the concept *unit* in Sec. 4. Next, two ways of organizing units and quantities are described, in Sec. 5 in terms of systems of units and in Sec. 6 in terms of application or disciplinary categories. In our schematic descriptions of the ontology we use the Rumbaugh OMT notation. The superscript “s” (^{s}) of an attribute indicates the multi-cardinality of that attribute. Boxes with underlined titles indicate instances. UnitDim can be downloaded from http://www.atoapps.nl/foodinformatics, Sec. *News* (no login is required) [7].

2 Overview

As already stated, in the past, several efforts have been made to design ontologies of units in some form or another. We have analyzed a selection of the most serious existing ontologies[1], including STEP [1], EngMath [2], MathML [3], OpenMath [4] and SCADA [5]. We have distinguished a number of quality aspects, ranging from consistency of terminology to questions like "Are units with prefixes distinguished from the elementary units without prefixes?'' , "Is it possible to convert units using this ontology?", "Are systems of units defined?" and "Are dimensions of quantities given?". Given these criteria, it appears that the existing descriptions are either incomplete, not presented in terms of standard ontology languages or only presented in terms of typical examples. These shortcomings made it impossible to build further on one of these approaches. However, we do base our ontology on the paper standards that are avail-

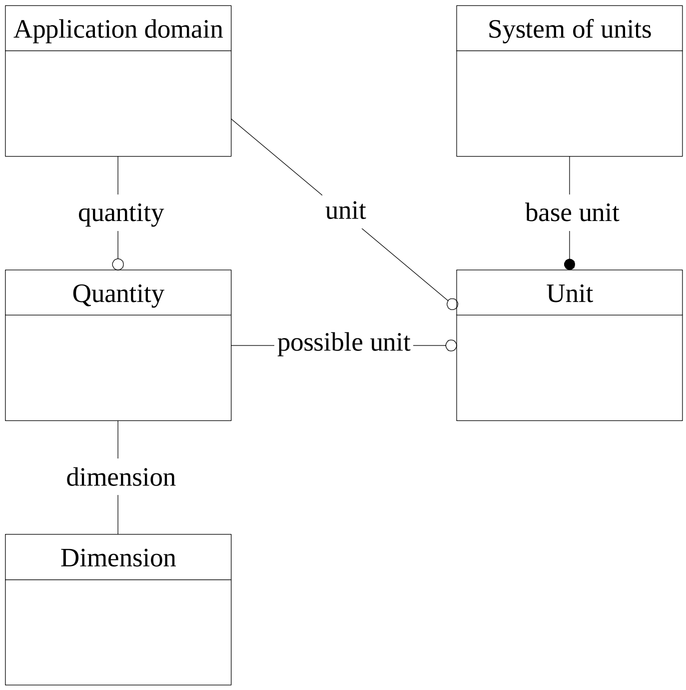

Fig. 1. Class diagram of the central concepts *quantity*, *unit*, *system of units*, *dimension* and *application domain* in UnitDim.

---

[1] This analysis will be reported in a separate publication.

able and widely accepted. It has been a challenge to explicate the freely formulated rules in these paper standards, often solely explained by means of examples.

To ensure that our ontology is actually serving a practical purpose, we have identified a number of "services" in e-science that it should be able to support:

- For a given quantity, provide useful units.
- For a given unit, provide quantities that can be expressed using that unit.
- Perform dimension checking on a set of equations
- Unit consistency checking.
- Convert from one unit to another.
- Provide meaningful quantities and units for a specific problem context.
- Give explanation and context of the quantities and units used.

Given these requirements it is clear that UnitDim cannot be only about physical units. Fig. 1 shows the basic concepts within the ontology and their relations. In order to obtain an applicable units ontology, we believe that the formalization of *quantities* is a precondition. In science, quantities are the central elements in terms of which knowledge is formulated. Subsequently, knowledge can be postulated in the form of statements involving quantities and their magnitudes. When we express the magnitude of a quantity, a *unit* is needed as a common reference. For many quantities more than one possible unit exists. The part of UnitDim that covers units and quantities already supports the first two services given above.

Next, the notion of *dimension* is needed. Each quantity refers to an abstract basis relative to a given *system of units*. The notion of dimension allows consistency checking in models and data, which is a powerful tool in science and engineering. Note that dimensions are not related to units directly. It is possible to add up yards and meters (with appropriate processing) but not meters and seconds. To support unit conversion a conversion factor within a specific system of units is required in addition to the dimension.

The next service, providing meaningful quantities and units for a given application, requires the organization of quantities and units in categories. A first filter is already given by the selection of a system of units. A second selection criterion is the type of application domain or discipline that typically requires a specific unit. For example, *mass* is a quantity related to the mechanics domain. Quantities and units may appear in more than one domain.

Finally, textual descriptions are required to explain quantities and units. In ModelDat we will also provide generic model statements that provide examples and explanations on how to use specific quantities.

3 Quantities, the building blocks of science

Objects and attributes are elementary concepts when representing the world. We claim that in science and engineering not *objects* but *attributes* are the primary concepts, since these can be related through theories and models. Two types of attributes can be distinguished: *qualities* and *quantities*. Qualities cannot be expressed on a numerical scale but need ordinal scales. A quantity of an object refers to “the extension of that object in a certain abstract direction”. A quantity is a measurable or, more generally, quantifiable physical property of that object. Examples of quantities are length, mass and time.

Fig. 2 displays some quantities in their class hierarchy. A quantity has a (textual) definition. For example, pressure may be defined as the force exerted over a surface divided by its area. A quantity refers to an object. The quantity *length of my table* refers to the object *my table* and is an instance of the class *length*, which is a subclass of *quantity*. In scientific and technical documents the object of a quantity is often left unspecified as it is assumed to be implied by the context. However, this can easily become a cause of misinterpretation.

The *symbol* of a quantity is a shorthand representation. For many quantities more or less standard symbols are used, but since these conventions are not always adhered to by individuals confusion can easily occur. The symbol of a quantity, printed in italic

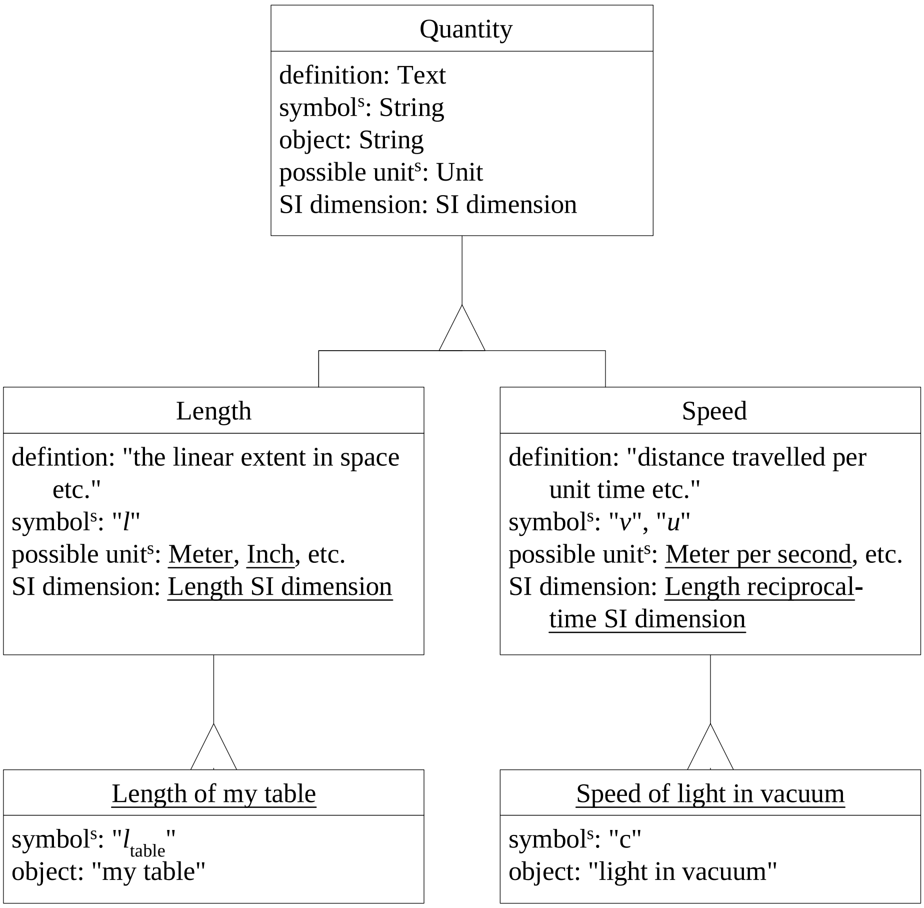

Fig. 2. Class diagram of the concept *quantity*.

type, should be single Latin or Greek letters, The two-letter symbols of dimensionless numbers are an exception to this rule [8]. Symbols used as subscripts and superscripts are, generally, roman if descriptive of nature. Symbols of matrix quantities are also italic. However, quantity vectors are boldface italic. Tensors are printed in sans-serif bold italic font type [8].

A few naming conventions apply for quantities, such as the use of the terms *molar* in case of quantities divided by the amount of substance and *specific* for quantities divided by mass.

The attribute *possible unit* contains all possible alternative units that are applicable for this quantity. The quantity *length* for example has possible units meter, yard, inch, foot, and so forth - in fact, exceptionally many alternatives; most quantities have less. *SI dimension* defines the quantity along the dimensional basis of the SI system of units. *CGS dimension* would do this for the CGS system, et cetera for other systems of units. In Sec. 5.2 dimensions are discussed in further detail.

4 Units of measurement for quantities

Units define reference standards that express the (quantified) extension along a quantity's dimension. Once the standard measure has been established, a quantity can be

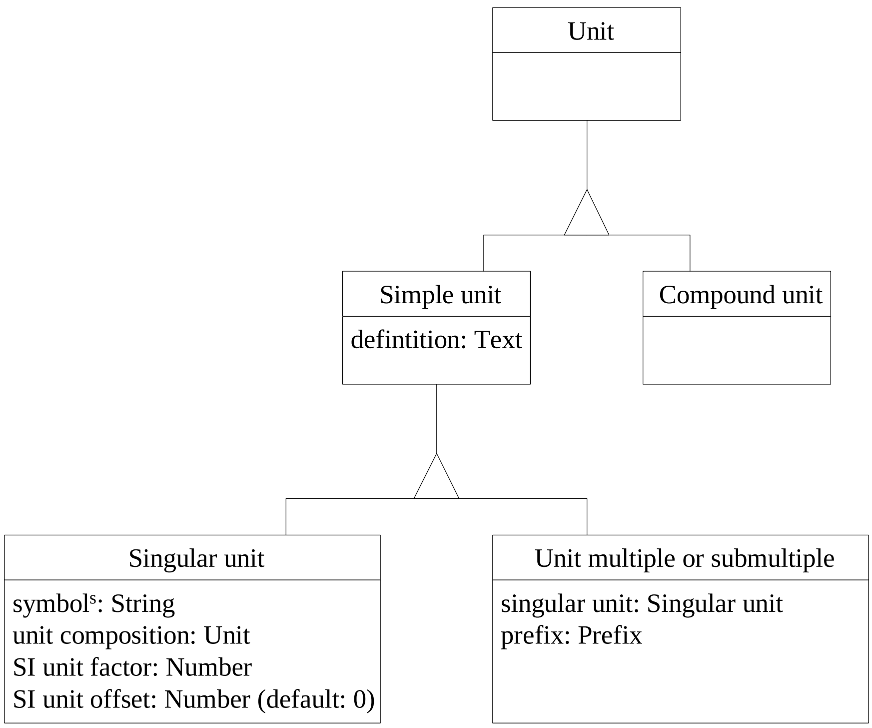

Fig. 3. Class diagram of the concept *unit*.

assigned a numerical value and becomes amenable to numerical computation. Hence, units are needed when numerical values are used.

We distinguish *simple units* from *compound units*. Examples of simple units are *meter*, *hertz* and *kilogram*. Simple units may have a definition. Compound units are (mathematical) composites of simple units or other compound units. Examples of compound units are *cubic meter*, *pascal second* and *newton per square meter*. Simple units in turn are either *singular units* (meter, hertz) or so-called *multiples and submultiples of units* (kilogram, decibel).

4.1 Singular units

Singular units are units that have a special name and (often) a symbol. *Meter, gram* and *kelvin* are examples of singular units. The full unit name is generally not capitalized, not even if they are named after people. Unit symbols are also not capitalized, in general, only if named after a person, in which case the symbol will *start* with a capital. Unit symbols are written in roman (upright) font type. Unit symbols are not to be pluralized and not to be followed by a period, except, of course, at the end of a sentence. Abbreviations other than the given symbols are not permitted [9].

Some singular units remain implicit, i.e., they do not show. This typically occurs when the associated quantities are dimensionless. An example is the unit for pH. The expression $pH = 7$ actually implies that $pH = 7$ pH-units.

The conversion factor of a unit to SI units is given by the attributes *SI unit factor* and *SI unit offset*, combined with the dimension of the associated quantity. The SI unit offset is important in case absolute temperatures and times (years) are converted from

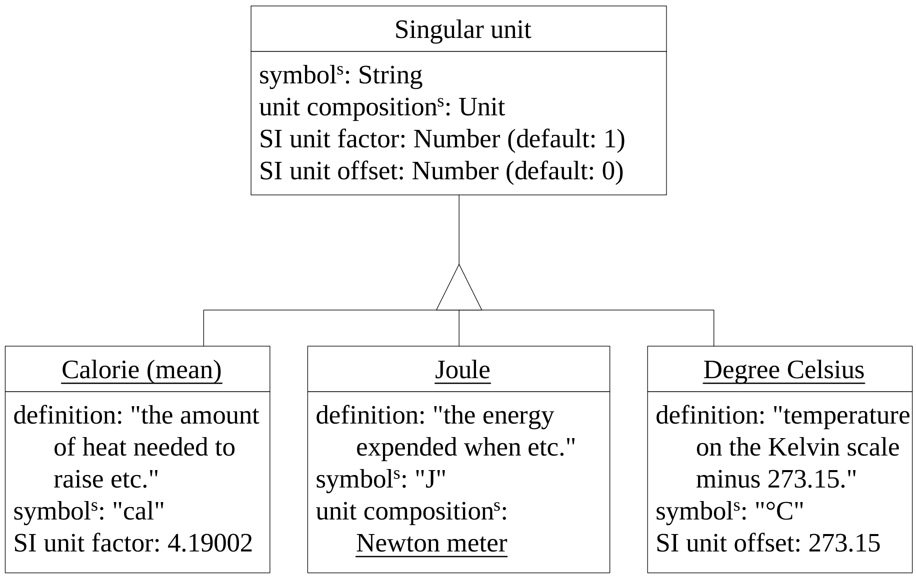

Fig. 4. Class diagram of the concept *singular unit*.

the one scale to the other. Sometimes it is meaningful to give the composition of a unit in terms of other units, by means of the attribute *unit composition*^{s}. An example is the *weber*, which is equivalent to *volt second*, a unit that is sometimes preferred when expressing magnetic flux.

4.2 Multiples and submultiples of units

In order to avoid very small or very large numerical values in practice, prefixes are used to form decimal and binary multiples and submultiples [9] of units. SI prefixes, representing powers of ten, are widely known. For example, attachment of the SI prefix *kilo* to a unit expresses a thousandfold of that unit. Although being called *SI* prefixes, these prefixes are also used *outside* the SI system of units (e.g. the decibel employs the prefix deci, but the bel is not an SI unit).

In addition to decimal prefixes, binary prefixes were introduced by the International Electrotechnical Commission (IEC), to offer a format preventing erroneous use of the SI prefixes in computer science [10]. For example the prefix *kilo* is commonly used to indicate 1024, instead of 1000 since $2^{10} = 1024 \approx 1000$. To prevent this abuse, the binary prefix *kibi* is introduced, representing exactly this factor 1024.

It is not permitted to use more than one prefix together with a unit. When printing, no space or hyphen may be used between prefix and unit. Not all combinations of prefixes and units are in common use. For example, only prefixes representing negative powers of ten (centi, milli, etc.) are used with the liter; only prefixes representing

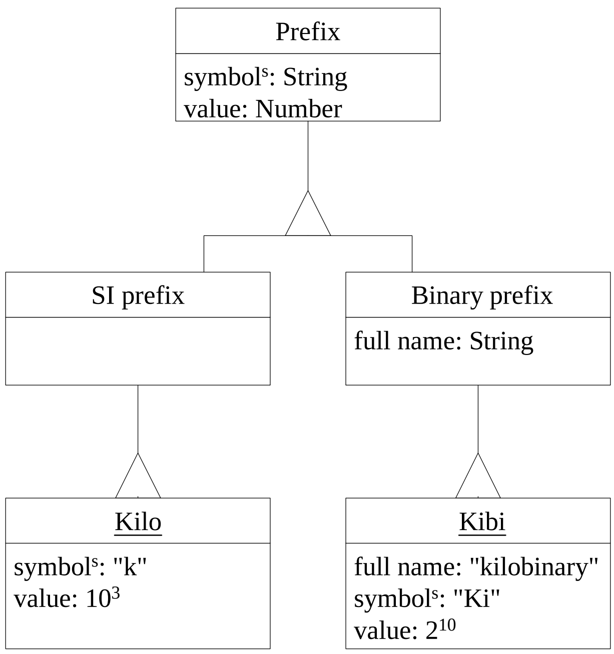

Fig. 5. Class diagram of the concept *prefix*.

positive powers of ten (kilo, mega, etc.) are used in combination with the metric ton (tonne). Among other units, prefixes are not used with the time-units *hour*, *minute*, etc. and angle-units *degree*, *minute*, etc. [9]

Like singular units, multiples can have a definition. *Kilogram* is an example of such a unit, being a base unit in SI and therefore having a definition. However, the definition of most multiples will depend on the prefix used and the definition of the singular unit prefixed.

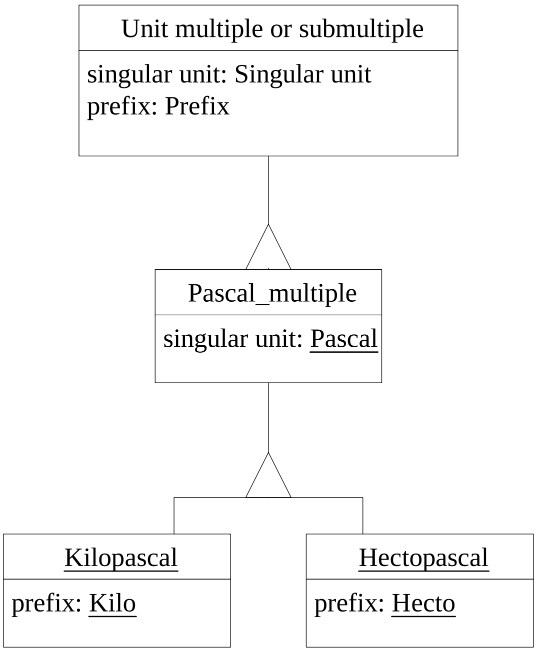

Fig. 6. Class diagram of the concept *unit multiple or submultiple*.

4.3 Compound units

Compound units are formed by multiplying or dividing one ore more simple units (singular or multiple). *Joule per second* is an example of such a unit. Here, the joule is divided by the second. A singular unit can refer to an equivalent compound unit through the attribute *unit composition*. For *watt,* for example, this is *joule per second*. This attribute provides alternatives that are used for their explanatory value in a particular application.

Basic unit operations are division, exponentiation and multiplication. More complex operations are division-exponentiation (a division with an exponentiation in its denominator) and division-multiplication-exponentiation (a division with a multiplication in its denominator, where one of the terms is an exponentiation). Explicit definition of this set of complex operations is needed to prevent for meaningless combinations of units in more complex combinations from appearing as separate compound units in the ontology. An example is *second squared*, which occurs in many more complex unit expressions, but makes no sense in a physical way.

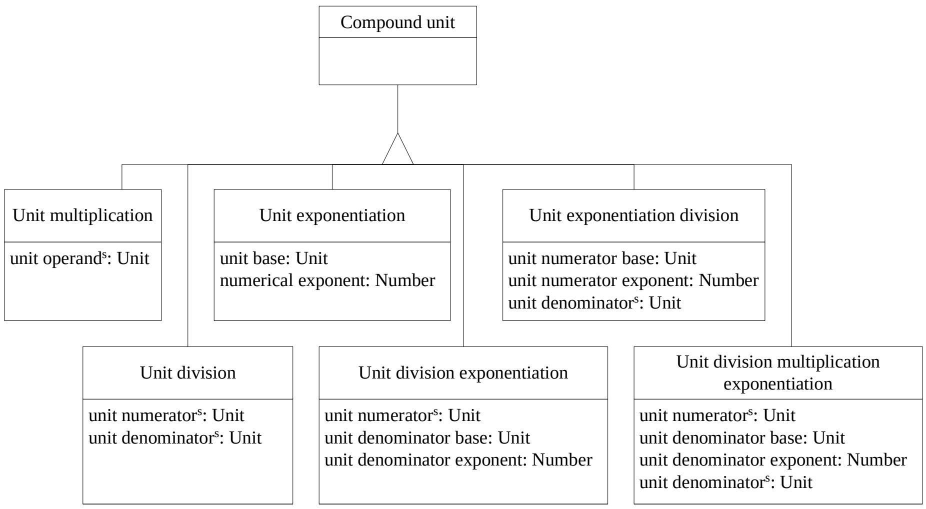

Fig. 7. Class diagram of the concept *compound unit*.

The example unit division *joule per mol kelvin* has unit numerator *joule* and unit denominator^s *mole* and *kelvin*. An example of a unit division exponentiation is *watt per square meter*, with unit numerator *watt*, unit denominator base *meter* and unit denominator exponent 2.

Unit multiplications are indicated using a centered dot or a space. Unit names and symbols are not used together. In unit multiplications and divisions it is preferable to have only one unit with a prefix, in order to avoid confusion. When printing the full names of unit multiplications only a space is used to separate the individual unit names, no hyphen or whatsoever. With unit divisions the word *per* is used, such as in *joule per second*. To avoid ambiguity, only one solidus (“division stroke”) may be used in a unit division, such as in J/(K kg). Parentheses are used when more than one unit appear in the denominator of the unit division. In general, the terms *squared*, *cubic*, *to the fourth power*, etc. are placed after the unit name in case of unit exponentiations. It is unacceptable to use mathematical operator symbols in combination with full unit names. Unit symbols should be used in that case instead [9].

5 Systems of units and dimensions

5.1 Systems of units

Through the centuries, an enormous number of units have been proposed. Many countries and regions had and still have their own units or versions of units. This has caused severe problems in science, but also in economy, trade and everyday life. People have and had problems understanding each other due to lack of standardization of units. In an effort to organize units in a proper way, they have been grouped in terms of *systems of units*. SI is the most comprehensive and best normalized example of such a system. Within a system of units the relations between the units it defines are postulated, in the form of mathematical expressions, often involving conversion factors. All units are either *base units* or combinations of base units, so-called *derived units*. Derived units can either be singular (*e.g., newton* in SI) or compound (*e.g., newton meter*). SI and CGS are *coherent* systems [11], meaning that derived units are related by a factor of 1 to the base units and all scaling factors are expressed in terms of prefixes. Other systems, as for example the British system of units, do not have this property. For example, the derived unit *inch* relates by a factor of 0.0254 (by definition) to the base unit meter. In Fig. 1 systems of units are related to units by means of the n-ary property *base unit* or *derived unit*.

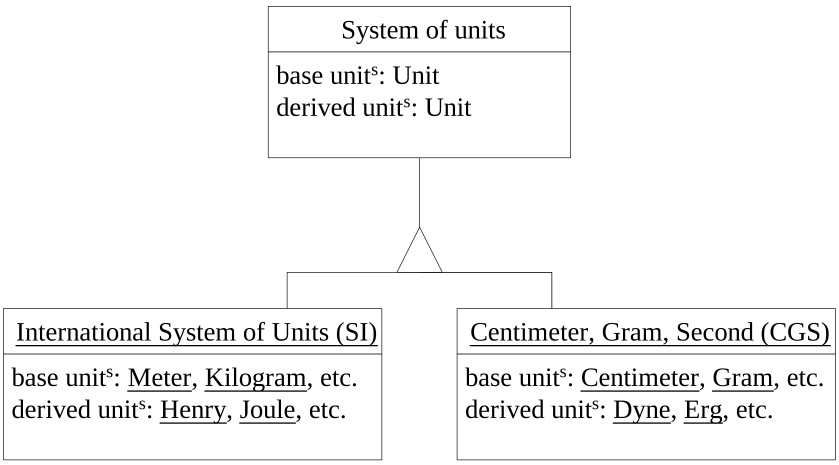

Fig. 8. Class diagram of the concept *system of units*.

Different systems of units that exist in the world generally have different sets of base units. For instance, in SI, the unit of mass, *viz.*, *kilogram* is a base unit, but in the British system of units, the *slug*, is a derived unit. In that system the *pound* (also referred to as pound-force) is a base unit.

Finally it is important to note that some units are not formally included in any system of units, but still can be expressed in terms of the base units of one or more of the systems. The torr is such a unit.

5.2 Dimensions

It is common practice in science and engineering to verify the consistency of mathematical expressions by “checking the units”. Inconsistent use of units points to sloppy or even erroneous modeling. However, in fact this type of model verification should not be bases unit checking, but on the analysis of *dimensions*. The dimension of a quantity points out in terms of which base quantities, for a given system of units, that quantity can be expressed, using exponents to express the mathematical relation. For example, in SI, the dimension of *force* has length exponent 1, mass exponent 1, time exponent –2, electric current exponent 0, temperature exponent 0, amount of substance exponent 0 and luminous intensity exponent 0.

Like base quantities, dimensions are defined within a system of units. SI uses *mass* in its dimensions, whereas the British system has *force* as an elementary dimension. Although strictly spoken a dimension is an abstraction of a (compound) quantity, in practice there is no real difference between the expression of a quantity in terms of base quantities or in terms of its dimension.

A quantity has no dimension if it cannot be expressed in terms of the base quantities of the chosen system of units. In practice many variables exist that are ad-hoc and

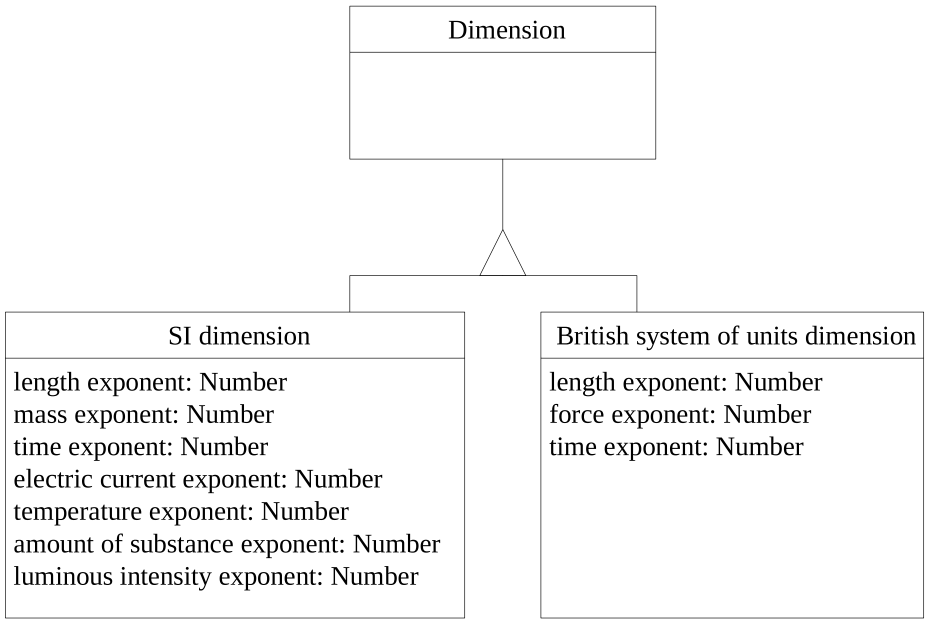

Fig. 9. Class diagram of the concept *dimension*.

cannot be expressed in terms of base quantities of any system of units. Moreover, it is not certain if it will ever be possible to express any quantity that comes up in practice in the given base quantities. It might even be necessary to define new base quantities accordingly. Relating new quantities to existing base quantities or defining new base quantities is a matter of increased scientific insight and consensus.

Returning to the issue of model verification using dimensions rather than units, we note that automated systems will be able to exploit this fact more what is common presently. Using UnitDim, it is perfectly well possible to add yards to meters or torrs to newtons per square meter, provided that numerical scaling is done using the proper conversion factors. The associated quantities are indeed compatible given the underlying dimensions.

6 Grouping quantities and units by their application domains

As stated in Sec. 2, one of the functions of an ontology of units and quantities in e-science is to suggest relevant quantities or units for a given problem context. For this purpose the concept of *application domain* is introduced. A common categorization is based on scientific disciplines such as mechanics, electromagnetics, fluid dynamics,

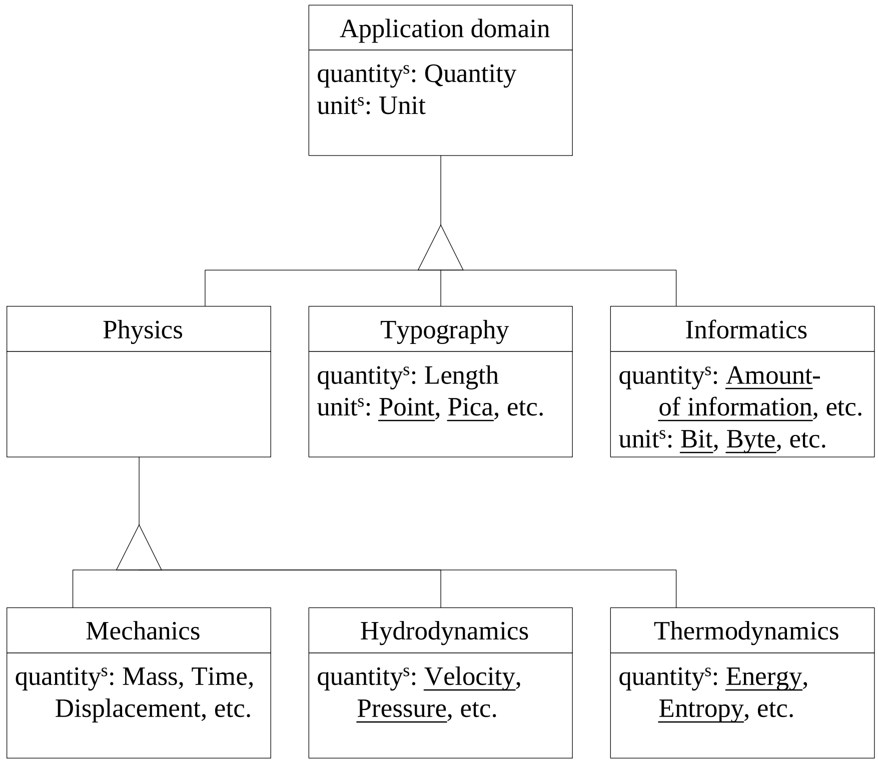

Fig. 10. Class diagram of the concept *application domain*.

thermodynamics, etc. Some quantities or units appear in more than one domain. Energy, for instance, occurs in all above mentioned domains. It should be noted here that these applications are not restricted to physics only. Other examples of application domains are economics, computer science and typography.

However, other ways to group quantities and units are also useful, as for example in something like *dimension groups* (e.g. space and time, dimensionless numbers). It appears to be necessary to specify both the quantities and the units for a specific domain. It is not sufficient to collect the units from the specified quantities this domain, since some units are not used in some domains. For example, in astronomy the point is not used as a unit of length.

7 Conclusions

In our daily business of performing industrial research projects we have run into the absence of a proper formal, digital representation of units and related concepts. After analyzing existing formalizations we concluded that these are either far from complete, not suitable for practical purposes or not formalized in terms of current representation standards. Having in mind the functions (services) that such an ontology should support we have constructed *UnitDim*. Although we do not claim completeness, we think that UnitDim covers most scientific and engineering applications. *UnitDim* presently contains more than 200 quantities and over 300 units and is implemented in OWL DL. Application domains covered are physics, chemistry, informatics, typography, etc. The ontology is proposed to be the basis for an international ontological standard and is offered for review to W3C.

Although seemingly straightforward, modeling units and measures proves to be an intricate task. Among others, the following design decisions were made:

- Quantities are the primary concepts and refer to dimensions. Nevertheless, UnitDim allows direct use of units, without referring to quantities, if desired.
- The application domains have been chosen following some standard reference works. However, in practice additional and more specific categories will have to be introduced.

Some issues that were not yet covered in UnitDim are:

- Order of magnitude information. Selection of appropriate units and prefixes for a given application could be guided by associating an order of magnitude to an application domain. For example, microbiological research typically deals with volumes in the range of μl to ml.
- In principle we are now able to establish a (large) number of rules for UnitDim. An obvious rule would be that in SI any unit can be combined with any prefix. Other rules that could be implemented are the printing style conventions of the several unit types and quantities.

We submit that many errors and misconceptions with units and quantities can be prevented once services and applications are directly available to scientists and engineers, based on a proper ontology of units and measures. *UnitDim* may prove to be such an ontology, in particular if it is enriched with input from practical applications.

UnitDim can be downloaded from http://www.atoapps.nl/foodinformatics, Sec. *News* (no login is required) [7].

Acknowledgements

This paper was written within the framework of the Virtual Lab e-Science (VL-e) project, subprogramme Food Informatics, financed by the Dutch Ministry of Education, Culture and Science and the Dutch Ministry of Agriculture.

References

1. ISO 10303 Standard for the Representation and Exchange of Product Data (STEP). http://www.iso.org (2000)
2. Gruber, T.R., Olsen, G.R.: An ontology for engineering mathematics. In: Doyle, J., Torasso, P., & Sandewall, E. (eds): Proceedings Fourth International Conference on Principles of Knowledge Representation and Reasoning. Morgan Kaufmann, San Mateo, CA (1994) 258-269
3. Harder, D.W., Devitt, S.: Units in MathML. http://www.w3.org/TR/mathml-units (2003)
4. Davenport, J.H., Naylor, W.A.: Units and dimensions in OpenMath. http://www.openmath.-org/cocoon/openmath/documents/Units.pdf (2003)
5. Lead, D., Schröder, A.: Scada on web. RDF vocabulary for physical properties, quantities and units. http://www.scadaonweb.com/publications/NOTE-units/2002-08-05/NOTE-units.-html (2002)
6. W3C: Web Ontology Language (OWL). http://www.w3.org/2004/OWL (2004)
7. Rijgersberg, H., Top, J.: UnitDim. http://www.atoapps.nl/foodinformatics. Sec. News (2004)
8. Cohen, E.R., Giacomo, P.: Symbols, units, nomenclature and fundamental constants in physics. 1987 revision. Document I.U.P.A.P.-25 (SUNAMCO 87-1). International Union of Pure and Applied Physics, SUNAMCO Commission (1987)
9. Taylor, B.N.: Guide for the use of the International System of Units (SI). 1995 edn. U.S. Government Printing Office Washington, DC 20402 (1995)
10. The NIST Reference on Constants, Units, and Uncertainty. International System of Units (SI). Prefixes for binary multiples. http://physics.nist.gov/cuu/Units/binary.html (2004)
11. Sizes. Index to units and systems of units. http://www.sizes.com/units/coherent.htm (2004)