Use of a Quantitative Research Ontology in e-Science

Hajo Rijgersberg[1], **Jan Top**[12], **Marcel Meinders**[13]

[1]Agrotechnology & Food Sciences Group, Wageningen University and Research Centre,
P.O. Box 17, NL-6700 AA Wageningen, The Netherlands
{Hajo.Rijgersberg, Jan.Top, Marcel.Meinders}@wur.nl
[2]Department of Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 1081A, NL-1081 HV Amsterdam, The Netherlands
[3]Top Institute Food and Nutrition, P.O. Box 557, NL-6700 AN Wageningen, The Netherlands

Abstract

This paper makes initial steps towards an ontology of quantitative research, in support of computer-aided scientific research. It identifies some key elements of the quantitative research process, outlines an ideal workflow, identifies further requirements and discusses how some of these aspects can be implemented for e-science. An implementation is sketched in which objects are related to the ontology and in which processes in the ontology can be initiated on objects. In addition, we report the application of an ontology of units of measure and related concepts, constituting a part of the concepts required in an ontology of quantitative research, in a unit and dimension consistency checking service, an example of a quantitative research service.

Introduction

Quantitative research is the scientific investigation of phenomena and their properties and relationships using quantitative concepts such as numbers, measurement scales, units of measure, mathematical operations, tables, graphs, etc. The objective of quantitative research is to develop and employ mathematical models, theories, and hypotheses about real-life phenomena. Quantitative research is widely used in many scientific disciplines, such as physics, biology, sociology, and psychology. The process of measurement is central to quantitative research as it provides the connection between empirical observation and mathematical expression of the quantitative relationships.

Information technology is intensively used in quantitative research, for making calculations, i.e., numerical operations on quantitative information, and storing quantitative information. In a world of ever-increasing scientific knowledge, the development of advanced services to scientific research is getting more and more important. A special field within computer science engages this subject: e-science.

One of the huge problems in quantitative research is the difficulty of reusing quantitative information. Generally, this kind of information (experimental data, mathematical equations, programming code, data files, graphs, etc.) is difficult to find, interpret, and execute (Top 2003). For example, the desired information can not be found because the quantities that are searched for are not mentioned as such in the sources, it is not clear which units of measure belong to the numbers, or a model can not be executed because it is not in a suitable format.

An important underlying problem is the lack of suitable quantitative vocabulary in information systems (Top 2003). It has been argued that conventional computer languages are not suitable for specifying scientific knowledge (Krishnamurthy and Smith 1994). Models are usually expressed using programming code, which represents an impoverished version of the scientific information (Keller and Dungan 1999). Data are specified in text files and spreadsheets, in relatively free formats, or in databases, in rigid formats. Describing information generally remains at the level of informal comments or journal papers which are only loosely coupled with the quantitative information, requiring due consideration by authors and readers of these models and data.

As far as we are aware, no extensive research has been done into requirements to quantitative research vocabulary. In this paper, we investigate these requirements and make initial steps towards an ontology which realizes this vocabulary. We identify some key elements of the quantitative research process, outline an ideal workflow and discuss how some of these aspects can be implemented for e-science. An implementation is sketched in which objects are related to the ontology and in which processes in the ontology can be initiated on objects. In addition, we report on the application of an ontology of units of measure and related concepts, such as quantities and dimensions, in a unit and dimension consistency checking service. Units of measure and related concepts constitute a part of the concepts that are required in an ontology of quantitative research.

We illustrate our argument with a study of creaminess of mayonnaises and custards. Concerns about obesity has led to increased interest in reducing fat or oil contents of foods. This is difficult as oil contents plays an important role in the perceived creaminess of many products (De Wijk and Prinz 2007). Therefore, the aim of this study was to better control the creaminess in such products. To this end, first creaminess had to be understood, on the basis of panel research, and subsequently relations with material properties and parameters that affect creaminess, such as rheological/mechanical properties under deformation such as viscosity and stress/shear moduli, had to be found, determined by instrumental measurements.

Overall Structure of the Quantitative Research Process

Quantitative research is generally approached using scientific methods and is considered to follow a structural, often iterative process whereby evidence is evaluated, theories and hypotheses are refined, technical advances are made, and so on. Figure 1 shows an overall structure of quantitative research, based on e.g. Gauch (2003) and Langley (2000). First (1), a research question is formulated, based on existing knowledge reported in text and. A research question is usually an open statement, which will be further specified by hypotheses. The research question in our research example was “Which factors control sensory creaminess of mayonnaise?”. In (2), qualitative concepts are defined, such as mayonnaise, different kinds of mayonnaises, e.g., commercial mayonnaises and specially prepared mayonnaises, and ingredients like oil and egg yolk. (3) Quantitative concepts are defined in order to be able to measure the studied phenomena and to quantify the relations between them. Examples of quantitative concepts are oil content, creaminess, and viscosity. Next (4), the hypotheses are formulated. One hypothesis in our research example was: “Fat controls sensory creaminess”. (5) The studied phenomena are modeled and available data and models are collected from literature. The aim of any research project is to support or reject a hypothesis. This is generally done by deriving a hypothetical “fact” from the hypothesis (6), such as in our research example: “Fat controls the sensory creaminess of these six mayonnaise samples”. Subsequently such a fact is compared with available or newly obtained data or models. For this purpose, often experiments are performed (7). The studied phenomenon is carefully realized, often using a laboratory (8). In our research example, different mayonnaise samples were prepared, with differences in oil content. Subsequently, the phenomena of interest are observed or measured (9). The creaminess of the mayonnaise samples was tested using a trained sensory panel. Next (10), the obtained data are processed, using statistical methods or model simulations. Creaminess was related to oil content using principal component analysis, a mathematical method that transforms a number of possibly correlated variables into a (smaller) number of uncorrelated variables. Finally, the processed data or models are compared with the hypothesis, which is subsequently considered to be supported, rejected, or revised (11). In our research example, the hypothesis was considered to be supported based on the result that oil content explained more than 98% of the variance of creaminess. (12) The obtained new knowledge is generalized (the relation also applies to other foods), (13) described in text, and (14) reproduced by others.

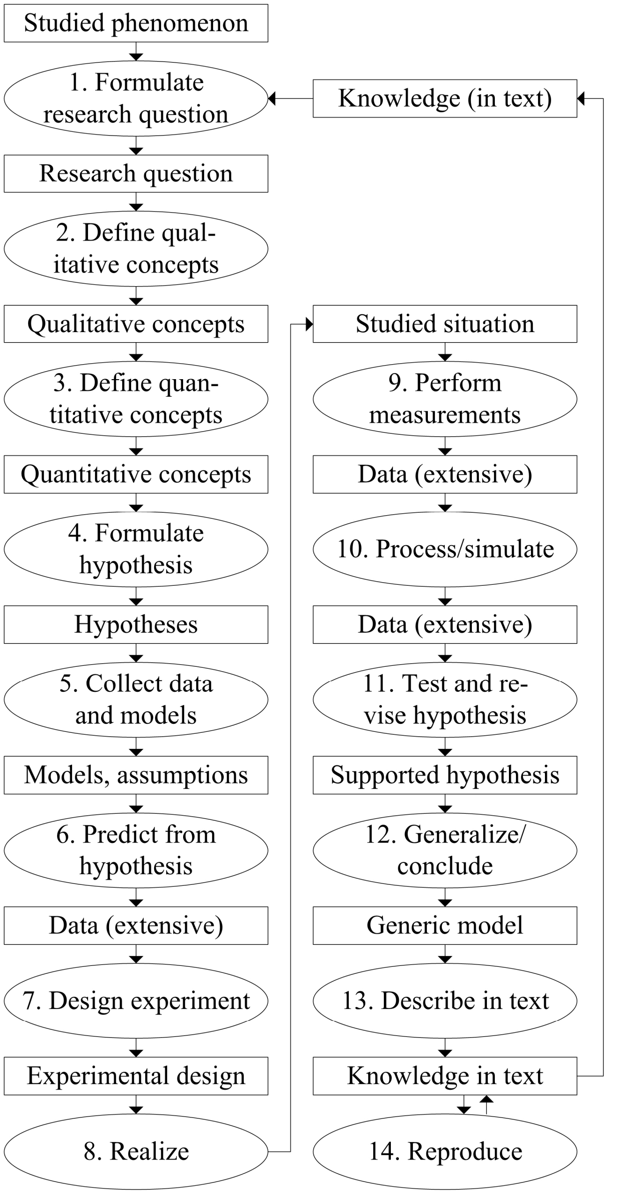


Figure 1. Overall structure of the quantitative research process. Ellipses indicate research steps, rectangles real-world phenomena and knowledge statements, and arrows input/output relations.

Not only phenomena from the real world can be subject to research, but also aspects of research itself. One such aspect is peer review. Peer review processes relate to performed research steps which are evaluated. Another aspect is method evaluation. New methods are evaluated by checking whether they have successfully led to valid knowledge statements or by comparison with (results obtained by) existing methods.

This framework is not the only way to categorize scientific activity, but it appears to have general applicability in discussing the current computer support of quantitative research processes and can be used as a basis for a sketch of quantitative e-science infrastructure, as we will show below. In daily practice, research steps will be omitted, repeated, performed in different orders, etc. The ultimate model of scientific research is still subject of debate in the disciplines epistemology and philosophy of science.

Computer Support of the Quantitative Research Process So Far

In this section we briefly discuss the computer support of quantitative research processes so far. Existing computer systems for the support of (quantitative) research processes or the storage of quantitative information include:

- Problem Solving Environments (PSE)
- Decision Support Systems (DSS)
- Data and Information Systems (DIS)
- Laboratory Information Management Systems (LIMS) and Laboratory Information Systems (LIS)

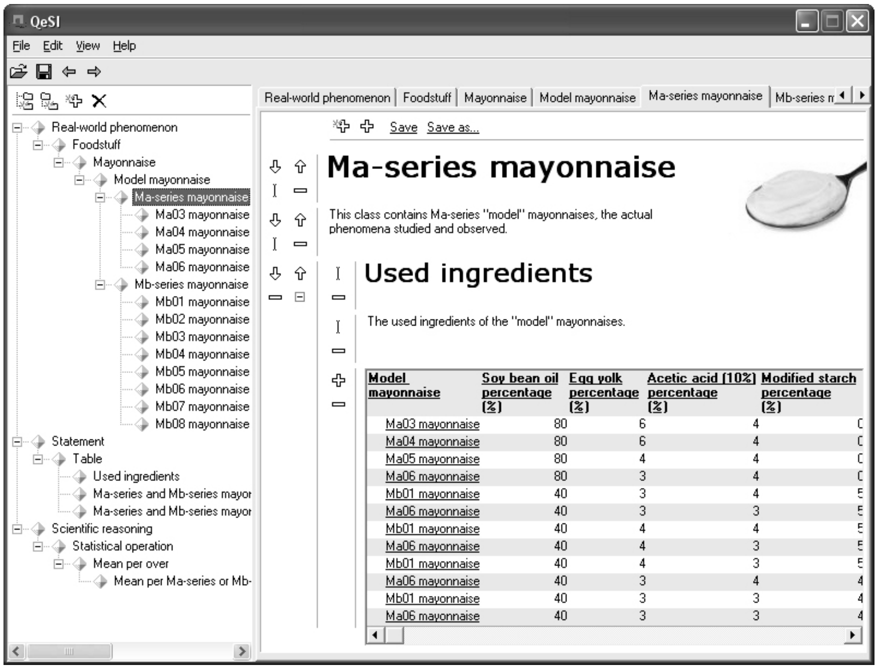
A PSE is a specialized computer system for solving special classes of problems. A DSS aids the process of decision making, where a decision is a choice between alternatives based on estimates of the values of those alternatives (Alter 1980, Finlay 1994). DISs are advanced computer systems that contain large amounts of data and calculation methods. In LIMSs and LISs, samples, instruments, laboratory users and functions such as invoicing, plate management and workflow automation are supported.

The list above is not exhaustive, but characterizes the diversity in computer support of research processes. The various systems cover different parts of the quantitative research process as shown in Figure 1. Since computer systems usually have their own vocabulary, restricted to their own scope, quantitative research is not specified in an integrated way. For example, it is not possible to indicate how a certain piece of knowledge was obtained, e.g., by measurement, analysis, or hypothesis formulation. This is considered to be a problem for the transparency of research in general and interpreting the validity of knowledge statements in particular.

The Semantic Web (W3C 2006) offers the possibility to define vocabulary externally from the computer systems that intend to use the vocabulary. This is accomplished by using languages such as XML, RDF and OWL (W3C 2007, W3C 2004). The use of standard formats and vocabulary is an important prerequisite for sharing vocabulary across multiple computer systems and platforms and, therefore, for reusing quantitative information.

Outline of a Quantitative e-Science Infrastructure

The previous section discussed that integrated support and vocabulary of quantitative research is currently lacking. In this section, we make initial steps towards an ontology of quantitative research, which realizes the desired



Ma-series mayonnaise

This class contains Ma-series "model" mayonnaises, the actual phenomena studied and observed.

Used ingredients

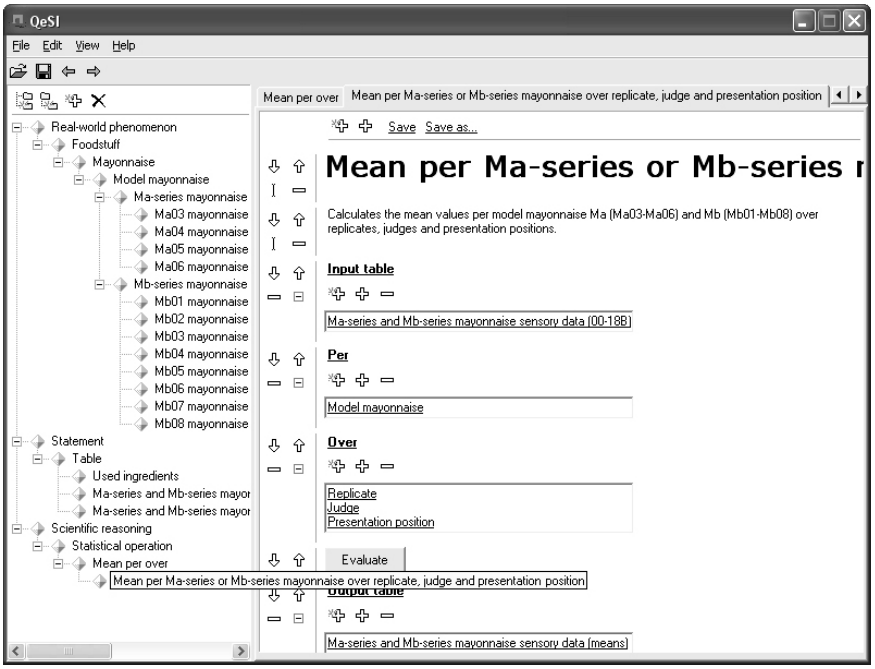
The used ingredients of the "model" mayonnaises.

Model mayonnaise	Soy bean oil percentage (%)	Egg yolk percentage (%)	Acetic acid (10%) percentage (%)	Modified starch percentage (%)
Ma03 mayonnaise	80	6	4	0
Ma04 mayonnaise	80	6	4	0
Ma05 mayonnaise	80	4	4	0
Ma06 mayonnaise	80	3	4	0
Mb01 mayonnaise	40	3	4	5
Ma06 mayonnaise	40	3	3	5
Mb01 mayonnaise	40	4	4	5
Ma06 mayonnaise	40	4	3	5
Mb01 mayonnaise	40	4	3	5
Ma06 mayonnaise	40	3	4	4
Mb01 mayonnaise	40	3	3	4
Ma06 mayonnaise	40	3	3	4

Figure 2. Sketch of a quantitative e-science implementation. A real-world phenomenon, “Ma-series mayonnaise”, is selected.

vocabulary, based on the framework shown in Figure 1. We sketch an implementation for e-science in which objects are related to the ontology and processes in the ontology can be initiated on objects. Figure 2 shows a screenshot of the implementation. In the left pane, (an excerpt of) a quantitative research taxonomy is shown. The root classes “Real-world phenomenon”, “Statement”, and “Scientific reasoning” can be identified. “Real-world phenomenon” defines the studied phenomena (qualitative concepts such as mayonnaises and ingredients) and quantitative, metrological aspects of these phenomena (e.g., oil content and viscosity). The samples that are used in the lab experiments are defined as instances of leaf classes of “Real-world phenomenon” such as “Ma-series mayonnaise” and “Mb-series mayonnaise”. Examples of such instances are “Ma03 mayonnaise” and “Ma04 mayonnaise”. The right pane shows details of the concept that is selected. For instance, for the class “Ma-series mayonnaise” the name, description and a table of ingredients that are used in preparing the samples are shown. The first two details are implemented as triples, the latter (the table of ingredients) is an instance of “Statement” which refers to the selected concept or subclasses or instances of the selected concept. Knowledge statements, typically, are data in the form of tables and models in the form of (lists of) equations. These equations contain all kinds of mathematical operations, such as multiplications, exponentiations, etc. In Figure 1, examples of statements are hypotheses, models, assumptions, and data, shown in the rectangles. The class “Scientific reasoning”, Figure 3, defines methods and operations that are performed on the real-world phenomena and statements. Some research steps in Figure 1 represent such scientific reasoning steps, such as “4. Formulate hypothesis”, “9. Perform measurements”, and “10. Process/simulate”. Figure 3 depicts an instance of a mean calculation method, “Mean per over”. The input properties of the method are restricted such that the operation can be performed (i.e., the calculation routine can be executed). The properties “Input table”, “Per” and “Over” are specified to, respectively, a table of measurements, the class “Model mayonnaise”, and some objects that we wish to calculate the mean over, namely “Replicate”, “Judge”, and “Presentation position”. Pushing on the “Evaluate” button generates a new output table. This new table is automatically stored – or rather defined – as a new statement in the ontology.

As subclasses of “Scientific reasoning” classes like “Measurement”, “Hypothesis formulation”, “Assumption formulation”, “Generalization”, “Induction”, “Deduction”, “Literature reference” and “Peer review” are defined in the ontology. To be able to rate quantitative statements originated from calculations, it is sometimes important to know which calculation routine of which software package was used. Therefore, an ontology of quantitative research should contain knowledge about systems, functions, services, etc. Usually, calculation methods in the ontology will be wrappers around external calculation algorithms. In addition, calculation algorithms can be represented in the ontology itself, as the ontology contains a mathematical branch.



Mean per Ma-series or Mb-series r

Calculates the mean values per model mayonnaise Ma (Ma03-Ma06) and Mb (Mb01-Mb08) over replicates, judges and presentation positions.

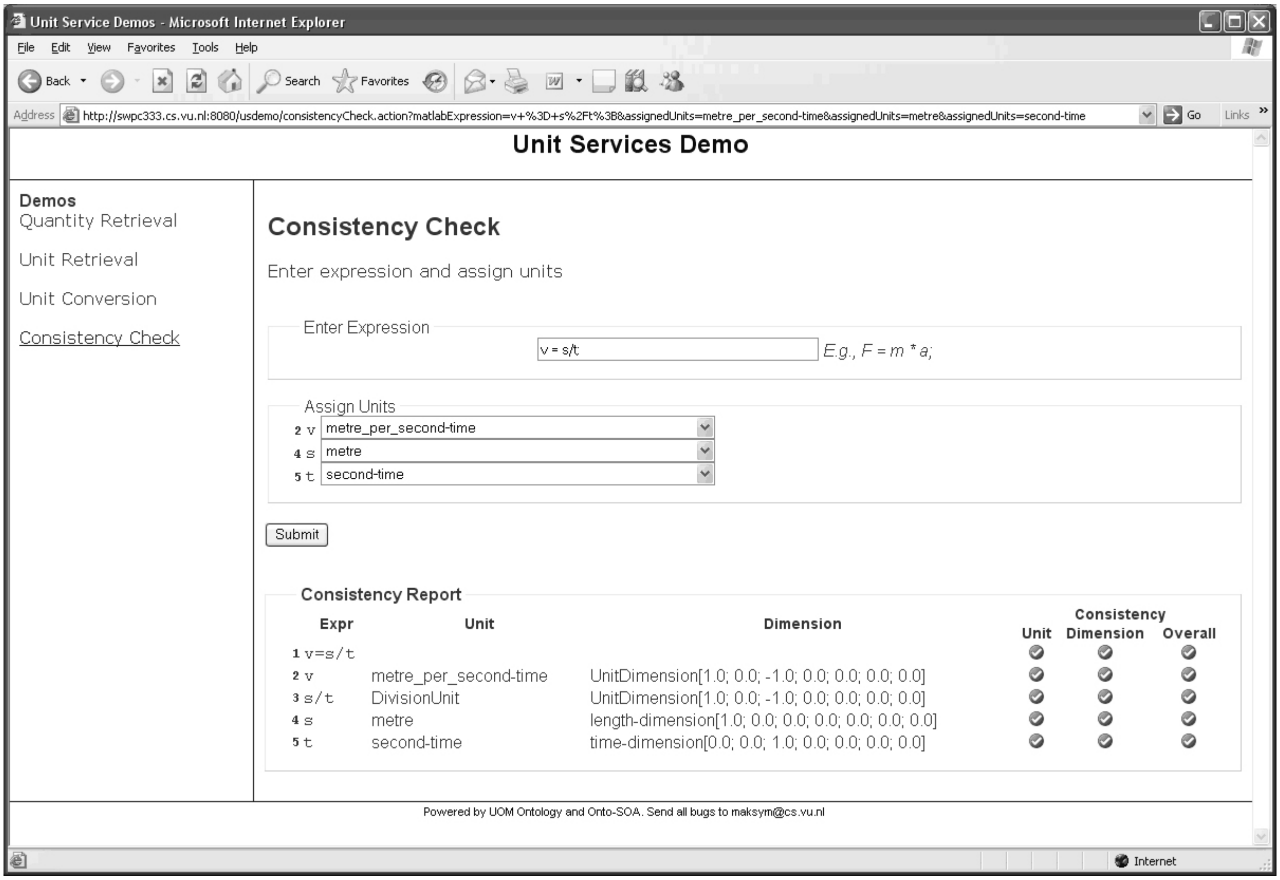
Input table
Ma-series and Mb-series mayonnaise sensory data (00-18B)

Per
Model mayonnaise

Over
Replicate
Judge
Presentation position

Output table
Ma-series and Mb-series mayonnaise sensory data (means)

Figure 3. A scientific reasoning operation, “Mean per Ma-series or Mb-series mayonnaise over replicate, judge and presentation position”, is selected.



http://swpc333.cs.vu.nl:8080/usdemo/consistencyCheck.action?matlabExpression=v+%3D+s%2Ft%3B&assignedUnits=metre_per_second-time&assignedUnits=metre&assignedUnits=second-time

Unit Services Demo

Consistency Check

Enter expression and assign units

Enter Expression: v = s/t   *E.g., F = m * a;*

Assign Units
2 v: metre_per_second-time
4 s: metre
5 t: second-time

Consistency Report

	Expr	Unit	Dimension	Consistency		
				Unit	Dimension	Overall
1	v=s/t			✓	✓	✓
2	v	metre_per_second-time	UnitDimension[1.0; 0.0; -1.0; 0.0; 0.0; 0.0; 0.0]	✓	✓	✓
3	s/t	DivisionUnit	UnitDimension[1.0; 0.0; -1.0; 0.0; 0.0; 0.0; 0.0]	✓	✓	✓
4	s	metre	length-dimension[1.0; 0.0; 0.0; 0.0; 0.0; 0.0; 0.0]	✓	✓	✓
5	t	second-time	time-dimension[0.0; 0.0; 1.0; 0.0; 0.0; 0.0; 0.0]	✓	✓	✓

Powered by UOM Ontology and Onto-SOA. Send all bugs to maksym@cs.vu.nl

Figure 4. Unit and dimension consistency checking service.

Application of an Ontology of Units of Measure and Related Concepts in a Unit and Dimension Consistency Checking Service

Units of measure and related concepts, such as quantities and dimensions, constitute a part of the concepts required in an ontology of quantitative research concepts. In earlier work (Rijgersberg and Top 2008), we constructed an ontology of units of measure and related concepts (OUM[1]). Unit and dimension consistency checking is an example of a quantitative research service. In our mayonnaise study for example, it is important to check the units of models that express relations between, for instance, viscosity and oil contents. Especially contents quantities can be expressed by various units of measure, such as percentages, g/ml, mol/m^3, etc.

Korotkiy and Top (2007) applied OUM in a unit and dimension consistency checking service. The concepts unit of measure, quantity, and dimension – among others defined in the ontology – were used for this purpose. In OUM, each quantity has a dimension and one or more units of measure, by means of which it can be expressed.

The user specifies a mathematical expression and for each variable assigns a unit of measure from the ontology. Subsequently, the service collects the quantities that are related to these units of measure from the ontology and, from these quantities, the related dimensions. Then, the service evaluates the expression according its unit and dimension consistency. The service appears to work well. Korotkiy and Top reported that the only problem they faced during their initial attempts, was the difficulty to access domain knowledge expressed using class restrictions in OWL. The relations between quantities and units of measure are expressed this way in the ontology.

Conclusion

In this paper, we made initial steps towards an ontology of quantitative research, in support of e-science. We identified some key elements of the quantitative research process, outlined a workflow, identified further requirements and discussed how some of these aspects can be implemented in e-science infrastructure. An implementation was sketched in which objects are related to the ontology and in which processes in the ontology can

---

[1] Ontology of Units of Measure and related concepts. The ontology can be freely downloaded from: http://www.atoapps.nl/foodinformatics, Sec. “News”.

be initiated on objects. In addition, we reported the application of an ontology of units of measure and related concepts, constituting a part of the concepts that are required in an ontology of quantitative research, in a unit and dimension consistency checking service, an example of a quantitative research service.

We concluded that integrated vocabulary of quantitative research processes is currently lacking in information systems. Such vocabulary is required for the advanced computer support of quantitative research. An ontology of quantitative research is a possible realization of this vocabulary. We showed that quantitative e-science can be structured around an ontology of quantitative research. The required vocabulary should include real-world phenomena, statements, and scientific reasoning. Especially the latter is important for the transparency of research in general and interpreting the validity of scientific knowledge statements in particular. A quantitative e-science infrastructure should be an intermediate between existing systems, functions and services. To this end, a quantitative research ontology should contain knowledge about systems, functions and services. In such an ontology, operations should be regarded as instances of methods, with their input properties restricted such that they can be carried out (e.g., calculation methods be executed).

Future work should focus on the implementation of quantitative vocabulary in an ontology and the application of such an ontology in advanced services which have to be developed. We are only beginning to design, implement and use ontologies of science in e-science. As more developers realize the need for collective and independent vocabulary and its use in research supporting systems, we will see a huge increase in advanced support of research processes.

Acknowledgements

This work was carried out within the context of the Virtual Laboratory for e-Science (VL-e) project (www.vl-e.nl), subprogram Food Informatics. This project is supported by a BSIK grant from the Dutch Ministry of Education, Culture and Science (OCW) and is part of the ICT innovation program of the Ministry of Economic Affairs (EZ). Additional funding was obtained from the Dutch Ministry of Agriculture (LNV). We would like to thank Jeen Broekstra, Roelfina Idema, Maksym Korotkiy, Suzan van Wezel and the anonymous reviewer gratefully for their valuable comments and reflections.

References

Alter, S.L. 1980. *Decision Support Systems: Current Practice and Continuing Challenges.* Reading, Mass.: Addison-Wesley Pub.

De Wijk, R.A., and Prinz, J.F. 2007. Fatty versus creamy sensations for custard desserts, white sauces, and mayonnaises. *Food Quality and Preference* 18: 641-650.

Finlay, P.N. 1994. *Introducing Decision Support Systems.* Oxford, UK: NCC Blackwell; Cambridge, Mass.: Blackwell Publishers.

Gauch, H.G. 2003. *Scientific Method in Practice*. Cambridge University Press.

Hey, T., and Trefethen, A. 2005. Cyberinfrastructure for e-Science. *Science* 308: 817-821, doi:10.1126/science.1110410.

Keller, R.M., and Dungan, J.L. 1999. Meta-Modeling: a Knowledge-Based Approach to Facilitating Process Model Construction and Reuse. *Ecological Modelling* 119: 89-116.

Korotkiy, M., and Top, J.L. 2007. Blackboard-style Service Composition with Onto⇔SOA. In IADIS International Conference WWW/Internet 2007. Villa Real, Portugal.

Krishnamurthy, M.V., and Smith, F.J. 1994. Integration of Scientific Data and Formulae in an Object-Oriented Knowledge-Based System. *Knowledge-Based Systems* 7(2).

Langley, P. 2000. The computational support of scientific discovery. *Int. J. Human-Computer Studies* 53: 393-410.

Rijgersberg, H., and Top, J.L. 2008. Ontologies of units of measure. Forthcoming.

Stevens, S.S. 1946. On the Theory of Scales of Measurement. *Science* 103.

Top, J.L. 2003. *Food Informatics. Kokkerellen met modellen.* Amsterdam: VU Boekhandel/Uitgeverij.

W3C. 2007. Resource Description Framework (RDF). World Wide Web Consortium (W3C). Available at: http://www.w3.org/RDF (accessed December 11, 2007).

W3C. 2006. Semantic Web. World Wide Web Consortium (W3C). Available at: http://www.w3.org/2001/SW.

W3C. 2005. W3C World Wide Web Consortium. Leading the Web to Its Full Potential… World Wide Web Consortium (W3C). Available at: http://www.w3.org.

W3C. 2004. Web Ontology Language (OWL). World Wide Web Consortium (W3C). Available at: http://www.w3.org/2004/OWL.