# Outline of planned PhD thesis

Mats B. Andersson and Donald Broady
ILU, Uppsala University

"The evolution of Western civilization has implied the formation of a conception of what education is […]" (Ulf P. Lundgren, *Att organisera omvärlden*, 1979, p. 24). According to the modern conception education is accomplished by means of courses, with a beginning and an end. Education is often perceived as kind of a preventive measure, with the aim to prepare a person for problems that might arise much later in life.

However, with the new notions on life-long learning together with the prospects offered by new ICTs (Information and Communication Technologies) we can no longer take for granted that education is primarily realized by the means of courses. New opportunities arise in domains where information tends to be immediately available for example on the Internet and accessible when needed. We might anticipate an environment where information sources — available on-line from Internet or whatever comes after the Internet — might be apprehended as quasi-ubiquitous everywhere where you can carry your ICT-equipment, in much the same way as today's infrastructures for energy or water distribution are accessible by electricity outlets or water taps.

We will here outline a thesis work that will reflect some aspects of these emerging new conditions for information supply by studying design principles for digital information repositories intended to meet the immediate needs of teachers or students.

Sometimes the user needs the information in order to solve an acute problem. For example: In an e-mail from a friend, one of us (M.B.A.) was requested to translate the following sentence "Bonjour, j'ai mal à la tête". In a few minutes he managed to deliver an answer although he does not know French. He did, though, know how to use the translation service Babelfish offered at AltaVista's web site.

In other cases the user might need the information in connection with a more long-range learning process. The web site of the World Wide Web Consortium (W3C) is a good example of an information repository which functions as school for almost anyone who wishes to keep up-to-date with current Internet technology.

W3C's web site represents a new breed of information sources emerging at the Internet. Its content is well defined and well delimited, comprehensive and up-to-date, organized in a way which meets immediate knowledge needs, and managed under the auspices of an organization entrusted with authority in the field. Someone who wishes short information on which attributes that belong to a certain HTML-element might confidently expect to find a correct and authoritative answer at the W3C web site, as well as someone who needs more in-depth information on the development of addressing technology on the Internet or the underlying principles of the XML-standard. Thus, whether the aim of the visit is to find a quick solution to an acute detail problem or to scrutinize complicated matters the site provides information when needed.

Albeit the W3C web site is not primarily intended for educational purposes the existence of this new kind of repositories should inspire the development in educational applications. It might serve as a model for future development where courseware is no longer confined to a sequential syllabus structure but tends to be more or less ubiquitous and available for input into ongoing learning processes when, and only if, needed. One might as a contrast compare with a traditional elementary computer course where the student is taught to master ten functions of a database software package, and in the end will need only three of those functions (which means that seven where taught in vain) or an eleventh.

In educational settings these new kinds of information repositories might be called learning domains. A learning domain is a sufficiently comprehensive collection of modules (below referred to as learning objects), clearly-defined and demarcated, and thoroughly organized in order to support immediate learning needs.

The planned PhD thesis will explore design principles for such learning domains, and comprise studies of how they are used by teachers and students. The work will apply research methods from social science and epistemology as well as from information and communication technology.

## The problem

The main problem is: how is information to be organized in order to satisfy immediate knowledge needs?

Among sub-problems the following seem at the present stage of inquiry to be the most central.

**How to mould and divided knowledge?**
The content should be portable, i.e. as far as possible apply to international standards or de facto standards of markup language, graphic formats, linkage mechanisms etc. The user should not be troubled with conversion problems when incorporating information into, e.g., his or her own text archive or database.

Further, content must be modularized. Within the educational ICT community "learning objects" is a common term. Appropriate modularization offers several advantages. The author's or editor's work is facilitated if changes affect only a narrowly circumscribed component. The user — for example a teacher who compiles tailor made courseware — will be able to combine components drawn form various sources into a personalized collection.

**How to classify content?**
The learning objects should be described in a way that enables the author and the user to overview and navigate the collections. The planned PhD work does *not*, though, aim at developing automatic information retrieval (agents and AI-technology etc.). Instead the purpose is to achieve methods to organize and describe the modules in such a way that information retrieval can be undertaken by the user in accordance with his or hers own needs. From a technical point of view it is crucial to develop an appropriate use of metadata, i.e. the information about the information

**How to manage associative links?**
The objects of a certain learning domain might be grouped and classified in many ways. Depending on the person and the situation some learning objects might appear as more close and others as more distant. Someone who draws information from a learning domain on funguses might wish a classification that separates edible ones from poisonous. Someone else might wish to classify them by their growing locality.

Further, in addition to these kinds of classification principles there might be connections of a more associative kind: someone with an interest in mushroom soup might want to associative links to

information about cookery. The learning domain should allow for various classification and association principles, which should not be permanently brazed into the material but rather offered by an author, editor or information broker as an option to the user.

**How to diminish the gap between the users' expectations and the offered content?**
The less divergences between a users' expectations and the actual content of the learning domain, the better in terms of usability. A user needs support in order to get an instant overview of the nature of a certain learning domain and its anticipated uses. Once again appropriate metadata are needed. This problem calls for empirical studies that elucidate what teachers or students expect to find in a certain information repository compared to what is in fact to be found there. Such user studies are to be included in the PhD work. The results will contribute to the development of design principles for how, e.g., learning domains on the Internet should present themselves to the visitors.

**How might a learning domain satisfy both instant information support and long-term learning?**
Above we gave the example on how a person who does not know French might use ICT and appear as fluent in this language. This course of action might appear as cheating. However, the pocket calculator was during a period a forbidden facility in mathematics teaching and considered as a threat to genuine education. It is nowadays quite a legitimate tool. In the same way web based translation support might soon form an integral part of language learning, not as a substitute but as a complement to other teaching ingredients. The example illustrates a conceivable displacement of the notion of knowledge which probably will have a considerable impact on educational practices as well as give cause to clashes of opinion. Also this problem is well suited for empirical studies, i.e. on how teachers and students retrieve information from the Internet and integrate it into their own work. Today a well known problem, no doubt an example of cheating, is that students download papers from the Internet that they were supposed to write themselves. In the future there will arise much more complicated problems caused by the eroding frontier between "genuine" acquired knowledge and the uses of available on-line support.

# Relevant R&D fields

The planned PhD thesis will draw from several fields of research and development as:

**Educational research** on curriculum theory and didactics, concerning the selection, classification and organization of knowledge in educational settings.

**Traditions in social science and epistemology**, starting with Émile Durkheim's and Marcel Mauss' notion of primitive forms of classification.

**De facto standards for education in distributed environments**. Most important is probably the IMS project (Instructional Management Systems), which is developing a comprehensive set of specifications and prototype software for facilitating the growth and viability of distributed learning on the Internet.

**Learning objects**. A working group called LOMG (Learning object metadata group) has developed a framework LOMF (Learning object metadata framework).

**Markup languages**. Both the supervisor Broady and the PhD student-to-be Andersson has long experience in working with SGML (Standard Generalized Markup Language ISO 8879:1986). For the years to come the most intensive development will probably be related to the further development of the new SGML based standard XML (eXtensible Markup Language) intended for structured document interchange on the Internet.

**Metadata**, i.e. information about information, for example descriptions on the content of a document or the content or structure of a web site. The most important de facto standardization initiative in this area is RDF (Resource Description Framework) which is designed as a general metadata model for description of web resources. RDF employs XML syntax

**Structuring of information**. Of importance is the development of CKML (Conceptual Knowledge Markup Language), another XML application.

**Addressing mechanisms**. Http-based addressing is very difficult to manage in larger scale or when the document collections are large and complicated. When a file is moved from one server location to another the http links break. Therefore it is important to monitor the development of URN (Universal Resource Name) which give each resource a unique id, something like the ISBN for printed matter.

## Preliminary schedule for the thesis work

Year 1999: Studies into relevant research fields, development of a detailed research plan.
Year 2000: Development of design principles, construction of a prototype for a learning domain on the Internet. User studies on the uses of existing Internet resources.
Year 2001. Evaluation of the prototype, further user studies.
Year 2002. Synthesis, accomplishment of the thesis.

## References

For more information on the technologies and standards mentioned, cf:

http://wave.eecs.wsu.edu/WAVE/Ontologies/CKML/CKML-DTD.html
http://sdct-ntsrv2.ncsl.nist.gov/metadata/2NDmeet/Hill/LOMG/
http://www.imsproject.org/
http://www.w3.org/XML/
http://www.w3.org/Addressing/
http://www.w3.org/RDF/
http://babelfish.altavista.digital.com/

For an introduction to the SGML standard as well as downloadable SGML-related tools and prototypes, see http://www.nada.kth.se/~broady/diglit/