How do you know that this search result is worth clicking on? An approach to identify suitable metadata for E-Learning products

Radka Wittmer, Roland Streule, Damian Läge

Department of Psychology, University of Zurich, Switzerland

Key words: *semantic metadata, E-Learning products, information retrieval, dissemination, IEEE LOM*

Abstract:

This paper aims to assist instructional designers by presenting a strategy for identifying user-centered and context-based metadata for products developed within Higher Education Institutions (HEI). It argues that a minimal set of metadata is indispensable for orienting users and disseminating E-Learning products and stresses adjustments made to the existing metadata schemas in order to properly serve to a particular focus group. Furthermore, it contributes to the open discussion regarding which paradigms should be used to access and structure these products so as to assure their efficient retrieval. The described procedure is implemented in the project "Educational Landscape Psychology (edulap)" where a Performance Support System (PSS) for targeted retrieval is being developed.

1 Educational institutions and their educational metadata: An unsatisfactory situation

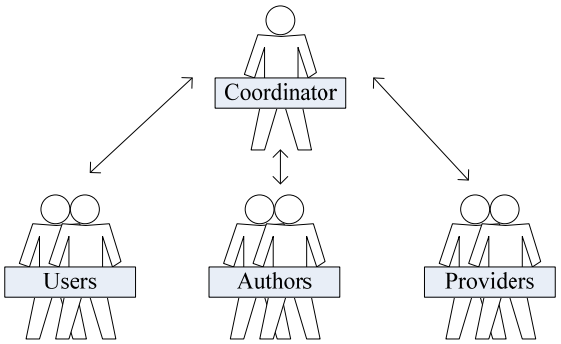
The majority of online educational products developed at different universities are commonly stored in the strategic Learning Management System (LMS) or FTP – Servers of these institutions. Simple full-text search engines are often used to retrieve these products [1]. To locate a satisfactory search result, it is therefore necessary to know the exact title. Otherwise, the users, i.e. students and lecturers, must search in an extensive list of results sorted by frequency of the entered key word(s).

Educational products are by-products of complex social and economic networks. An understanding of these networks and their relationships is needed in order to understand the products [2]. However, there are only a few or no available descriptors, i.e. metadata, [1] to explain these networks. The presence of metadata would make it easier for users to retrieve the products, and in addition to evaluate them for suitability based on their needs. Instead, currently used search engines have a negative impact on work satisfaction as well as on decision-making strategies and development of explorative and autonomous learning skills.

A brief survey in LMS Moodle [3], Ilias [4], Olat [5] and Dokeos [6] with a single sign-on web access available across Swiss universities [7] confirmed the above described situation. Only Ilias provides advanced search features. It has adopted the full version of 1484.12.1-2002 IEEE Standard for Learning Object Metadata (LOM) [8] therefore, it is possible to describe educational products in such a way that they can be integrated into the network. However, no educational products could be found using this advanced search, since they are not tagged with metadata. The absence of descriptors might be caused by lack of motivation to make the products public or by an inappropriate strategic decision regarding implementation of a general standard in English to a German speaking university without adapting it to the special needs of the user community.

2 User-centered and context-based metadata: Basis for dissemination and exchange of E-Learning products

Duff [2] recommends that users are involved in the development of metadata schemes to ensure that the correct metadata are identified at the appropriate level of specificity to meet users' needs. However, it is the interdisciplinary project network (Figure 1), which assures the development of a performance support system facilitating efficient retrieval. Therefore all the project members should be integrated equally into the decision-making process. They should ensure that the system, consisting of technical infrastructure, database structure, search engine, metadata editor and a graphical user interface, is to be developed in the best interests of each of them, i.e. the end user.



Initial role	Lecturers & students	Lecturers	Programmers
Activity within the system	Search & retrieve	Tag	Maintain
Task in the project	Select metadata & vocabulary	Select metadata & vocabulary	Evaluate & implement technical standards
System requirements	Intuitive presentation of results & their refinement	Tagging takes little time & products can be retrieved	Interoperability with other systems & reduction of complexity

Figure 1: Project network with flat hierarchy role distribution & characteristics of the target groups

Students generally prefer to work with online educational products which have a comprehensive didactic structure because of the transparent relationship between the whole and its parts. Commonly, they use a search engine to obtain answers to their questions concerning a certain topic or a task. In order to find the answers, they use content-keywords or purpose-keywords (e.g. repetition); search for materials developed by their teachers or for specifically recommended products. The transparent structure is not only crucial for the clarity of any product and understanding its message, it is also crucial for the presentation of search results in order to reach efficient decisions.

Lecturers already possess the didactic structure (in their classes) and therefore prefer products with a clear didactic function (e.g. prerequisites, presentation, post-test) or elements (e.g. figure, animation, film) of which these are consisted of. They search in order to implement the located search result into their classes or for inspiration on how others are presenting certain topics. Next to the content-keywords they search for names of their colleagues who work in the same field or for specific media types. So as the students do, also the lecturers expect that the search engine produces results which are intuitively understandable and allows efficient orientation among them.

Authors, who agree to make their products public, anticipate that these can be found by others. More often than not, authors want to invest as little time as possible with administration and

believe that it is sufficient to upload their products onto a repository. These repositories are often exclusively developed by *programmers who assure technical functionality*, i.e. products can be uploaded, described, stored, searched and retrieved. Existing metadata schemas, i.e. data models, approved through the technical community, are commonly chosen for the description of educational products. Adherence to such models assures the exchange of data with other technical systems and offers users a wider spectrum of educational products which they may access. However, such systems force authors to describe their products by using metadata which have only little relevance firstly, to the products themselves and secondly, to the context in which they should be used. Consequently, authors do not fill the metadata and thus students and lecturers can not obtain satisfactory search results. Naturally, such tools are given little or no attention and subsequently become unused.

3 Procedure: A particular approach

In the project edulap [9] partners from Swiss universities in Bern, Geneva, Lausanne, Zurich and Distance Learning University Foundation Switzerland are developing a PSS to be used for targeted retrieval of educational products. The goal is to present search results as orientation maps allowing users to rapidly see all relevant results [10], [11] and to provide immediate access to the E-Learning products.

Students, lecturers and authors of educational products are included in the identification of the appropriate metadata as well as in the development of the metadata editor and the search engine which is being evaluated for correctness of selected metadata (Figure 2). Programmers, with good knowledge of the target context, target users and their relation to technical solutions, are invited to contribute their expertise to the development.

In order to create the first metadata draft, metadata relevant for the HEI were selected from the existing LOM standard. Missing metadata, which are typical for HEI, were identified as well as search strategies of target users and the amount of time that authors are ready to invest. The obtained list of metadata is to be implemented and products described accordingly. Subsequently, users are to be asked to identify the most important metadata in terms of their needs (students and lecturers) and resources (authors). The goal of the second draft is to obtain a reduced number of metadata which must be implemented and tested in the latter stage. Next, the aim is to determine whether or not students and lecturers can evaluate the educational products for suitability to their own needs according to the provided metadata. It is also important to determine if authors can attach the required metadata within a reasonable time limit. Positive findings will lead to a final metadata set otherwise additional adjustments of the previous metadata will be carried out.

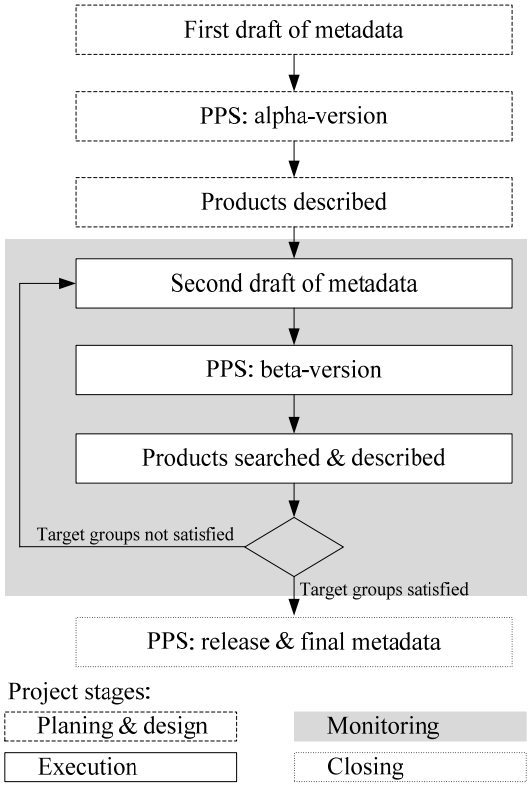


Figure 2: Process of identifying the user-relevant metadata and their direct implementation into the development of the search engine edulap

4 Development: Initial challenges

In the planning and design stage (Figure 1) the central issues (a-d) to be decided with respect to the context and needs of the target users are:

a) Which existing metadata schemas can be taken into consideration?
b) Which existing metadata schema should be selected?
c) Which metadata elements are missing from the selected schema or are redundant?
d) Are the values of the metadata elements clear and understandable?

Choosing wrong technology and a wrong metadata schema would have a cost exhausting impact on the project budget. Adding or deleting single metadata elements or editing vocabulary items are a subject of change in the monitoring stage, however major changes concerning technology solutions are, if possible, to be avoided.

Context predetermines the selection of possible schemas. Currently, there are dozens of metadata schemas each of which is either suitable for a different focus group or they have a more general character. For example the majority of the multimedia community uses the ISLE Meta Data Initiative (IMDI) [12] standard, which describes the multi-media and multi-modal language resources; or libraries use the Metadata Encoding and Transmission Standard (METS) [13] schema, which is a standard for encoding descriptive, administrative, and structural metadata regarding objects. Currently, the most well known schema in the educational community is the IEEE LOM, an international standard for describing learning objects, i.e. “any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning” [8].There are also the more general schemas like the one from Dublin Core Metadata Initiative[14], which is an international standard that supports a broad range of purposes. It goes without saying that each schema has a different technical realisation and a different set of metadata elements, characterising one focus group or another.

The purpose of the project is decisive in the selection of a particular schema. The selection of an appropriate schema, (not only in the field of education), is influenced by an obvious “tension between a more targeted approach, with much more limited appeal, versus a more generic approach, with broad application” [15]. Dublin Core is a very simple, generic set, with fewer than 20 tags and it is commonly used to describe resources across all domains. IEEE LOM is on the contrary a more targeted tag set, with more specific and detailed tags which are aimed at tagging objects that focus on learning. Its elements are designed to capture educational context and pedagogical information in addition to descriptive data.

Needs of the target group steer the modification of the selected schema. Not all metadata in a selected schema are relevant for a specific community of practice, moreover, there are always some metadata missing which are essential for providing context for the potential user. A strict adherence to a selected schema has thus not proved to serve the target community well. For example in the field of education the Dublin Core schema (DC-Simple with 15 elements) is considered to be “inappropriate for the great heterogeneity of content an institutional repository may hold…”. Bueno-de-la-Fuente et al. [16] provide evidence for “the need to use metadata schemas that provide more detail about specific domain resources”. A step in this direction is the development of application profiles. These are modifications of existing schemas consisting of reduced and/or extended metadata elements optimised for a particular context. Likewise, the vocabularies in the original schema may be supplemented with more appropriate values. However, in order to assure interoperability with other systems, the metadata must be stored in a form which can be exported as records of the original schema.

Values of individual elements must be intuitively understood by the target groups. Consistent use of language with metadata descriptions supports the consistent discovery of resources within one or several repositories. The primary tool for ensuring consistent language

usage is controlled vocabulary, including the use of thesauri. However; it happens very often that such thesauri are available in only one language. It would be irresponsible to implement a thesaurus in a foreign language and unacceptable to ask authors to describe their products using values that they do not understand. Such a request would not only have a negative impact on the quality of the product description but also on the obtained search results when trying to retrieve a particular product from the developed repository. Therefore it is necessary to provide translations in the mother tongue of the target users and test the understanding subsequently. Adherence to the original thesauri should be assured on the technical level.

To demonstrate the implementation of the above points, it was already stated, that the project edulap has selected the IEEE LOM schema to describe the products available at HEI in Switzerland. Nevertheless, the original LOM schema does not entirely reflect the nature of educational products available. It contains some metadata elements which are irrelevant to the given context or the user can obtain such information from other sources than the metadata. The process of amendment is presented in the three examples of metadata elements below.

I. LOM category: General; Element: Coverage

This element should inform the user about the extent of the content of the learning object ("e.g. 16th century France" [8]). The expert reviews with the authors of the educational products reviled that they tend to include this information into the abstract of their products, i.e. Element: Description. Therefore it was decided to modify this element and include standardised vocabulary of sub-disciplines of a particular field of expertise, depicting the division at Swiss universities. For example the field of psychology was further divided into social psychology, cognitive psychology, health psychology, emotion & motivation etc. The authors will have the possibility to allocate more values to one product, since students from other fields can also profit from one particular product.

II. LOM category: Technical; Element: Installation Remarks

This element should inform the user about how to install the learning object ("e.g. Unzip the zip file and launch index.html in your web browser" [8]). After conducting the expert reviews with programmers it was decided not to implement this element. Upon accessing an educational product, whose parts are used with an application which is not available on the computer of the potential user, the computer will notify the user and recommend installing the missing application first.

III. LOM category: Educational; Element: Learning resource type

LOM defines this element as a specific kind of learning object which is used for a specific learning purpose (orientation, revision etc.). The suggested values stem from the Oxford English Dictionary published in 1989. However, the variety of online learning resources has expanded greatly since then. Therefore, in the expert reviews with authors it was decided to take a subset of a current Learning Resource Type Vocabulary, published by National Science Digital Library (NSDL) [17] in 2006 and consisting of 43 items, representing the educational products available within Swiss universities.

5 Conclusion

The selection of metadata and a particular schema must result from a balance between the requirements of the academic community who accentuate the content and structure of online educational products, i.e. human-generated metadata, and those of programmers who support the accuracy of machine-generated metadata and adherence to a standardised data model. Moreover, a user friendly metadata editor, which would lead the authors through the tagging process, must be developed. Since authors are the key players in the success of the performance support system, the editor should make them aware of the positive consequences of adding

metadata for the search results. The process of entering metadata should be sequential, ordered from the most to the least important ones.

Implementation of any standardised model in a system which should serve a particular community of practice would act only as an elaborative norm model. This would indeed assure a functionality of the system; however, the authors and the target users will not see its relevance for their particular requirements. Consequently, the system would not be used. Attempts to reduce the complexity of norm models through the omission of certain metadata do not bring any satisfactory solution to this situation because the reduction of quantity has a negative effect on the quality of the described products and thus on correctness of the search result (a discrepancy between the expectations of the user and the presented search results). In such norm schemas, there are missing metadata depicting the context required both by authors of the educational products as well as by the target users. Therefore the success of a performance support system depends on cooperation of all member groups, readiness to reach compromises and mutual respect to the needs of all parties concerned.

References:

1. Sack, H. (2008) Semantisch unterstützte suche und Navigation in audiovisuellen Datenbeständen. In *3.tele-TASK Symposium* Hasso-Plattner-Institut, Potsdam, Germany.
2. Duff, W.M. (2001) Evaluating Metadata on a Metalevel. *Archival Science* **1**, 285-294.
3. LMS Moodle. http://moodle.org, [Retrieved: 23.04.2009].
4. LMS ILIAS. http://www.ilias.de, [Retrieved: 23.04.2009].
5. LMS OLAT. http://www.olat.org, [Retrieved: 23.04.2009].
6. LMS Dokeos. http://www.dokeos.com, [Retrieved: 23.04.2009].
7. SWITCH: Authentication and Authorization Infrastructure. http://www.switch.ch/aai/ [Retrieved: 23.04.2009].
8. Learning Technology Standards Committee: IEEE Standard for Learning Object Metadata 1484.12.2. http://ltsc.ieee.org/wg12, [Retrieved: 23.04.2009].
9. Educational Landscape Psychology. http://www.edulap.ch, [Retrieved: 23.04.2009].
10. Streule, R., Läge, D. (2008) Educational Landscapes: Mapping der elektronischen Ausbildungsangebote eines Faches mit Kognitiven Karten. In *Offener Bildungsraum Hochschule - Freiheiten und Notwendigkeiten* (S. Zauchner, P. Baumgartner, E. Blaschitz and A. Weissenbäck, eds), Waxmann, Münster 50-57.
11. Streule, R., Läge, D. (2009) Effective Orientation in the Field of Virtual Learning Resources. In *11th Congress of the Swiss Psychological Society*, University of Neuchâtel, Switzerland.
12. ISLE Meta Data Initiative. http://www.mpi.nl/IMDI, [Retrieved: 12.08.2009].
13. Metadata Encoding and Transmission Standard. http://www.loc.gov/standards/mets, [Retrieved: 12.08.2009].
14. The Dublin Core Metadata Initiative: Dublin Core Application Profile Guidelines. http://dublincore.org/usage/documents/profile-guidelines, [Retrieved: 23.04.2009].
15. Schatz, S.C. (2005) Unique Metadata Schemas: A Model for User-Centric Design of a Performance Support System. *Educational Technology Research and Development* **53**, 69-84.
16. Bueno-de-la-Fuente, G., Hernaacutendez-Peacuterez, T., Rodriacuteguez-Mateos, D., Meacutendez-Rodriacuteguez, E.M., Martiacuten-Galaacuten, B. (2009) Study on the Use of Metadata for Digital. Learning Objects in University Institutional Repositories (MODERI). *Cataloging & Classification Quarterly* **47**, 262-285.
17. NSDL Registry: NSDL Learning Resource Type Vocabulary. http://metadataregistry.org/concept/list/vocabulary_id/11.html, [Retrieved: 08.08.2009].

Author(s):

Radka Wittmer, M.Ed.
University of Zurich, Department of Psychology, General Psychology (Cognition),
Workgroup Applied Cognitive Psychology
Binzmühlestrasse 14 / 28
CH-8050 Zürich
r.wittmer@psychologie.uzh.ch

Roland Streule, Dr.
University of Zurich, Department of Psychology, General Psychology (Cognition),
Workgroup Applied Cognitive Psychology
Binzmühlestrasse 14 / 28
CH-8050 Zürich
r.streule@psychologie.uzh.ch

Damian Läge, Prof. Dr.
University of Zurich, Department of Psychology, General Psychology (Cognition),
Workgroup Applied Cognitive Psychology
Binzmühlestrasse 14 / 28
CH-8050 Zürich
d.laege@psychologie.uzh.ch