Semantic and Pedagogic Interoperability Mechanisms in the ARIADNE Educational Repository

E. Forte, F. Haenni, K. Warkentyne, *Swiss Federal Institute of Technology, Lausanne, Switzerland*

E. Duval, K. Cardinaels, E. Vervaet, K. Hendrikx, *Katholiek Universiteit Leuven, Belgium*

M. Wentland Forte, F. Simillion, *University of Lausanne, Switzerland*

Abstract *This paper reports on the principles underlying the semantic and pedagogic interoperability mechanisms built in the European Knowledge Pool System, developed by the European research project ARIADNE. This system, which is the central feature of ARIADNE, consists in a distributed repository of pedagogical documents (or learning objects) of diverse granularity, origin, content, type, language, etc., which are stored in view of their use (and reuse) in telematics-based training or teaching curricula. The learning objects are indexed, usually by faculty staff, according to the ARIADNE metadata set. The principles embodied in the indexation tool, which interacts directly with the repository, stem from a few theoretical ideas but foremost from empirical, pragmatic considerations, suggested by the context of actual use. They tentatively address the stringent demands for semantic and pedagogic interoperability implied by a context of rather wide cultural and linguistic diversity, as well as those stemming from the very nature of the domain application itself : education and training. Possible extensions to the educational metadata scheme developed by ARIADNE on these basis, may accommodate corporate training/information needs. These extensions are briefly discussed as a mean for enhancing 'semantic' interoperability between different (kinds of) corporations. Finally, the architecture of the ARIADNE system, which heavily relies on this educational metadata system, is briefly reviewed.*

1. Operational Issues : Pedagogical Indexation & Pedagogical Retrieval

The context of this work stems from the nature of the ARIADNE* Consortium and from the general goals pursued by this project: ARIADNE[1,2] currently brings together twenty four mostly academic partners from eight different European countries plus several corporate sponsors. The Knowledge Pool System[3] (KPS), which is the central feature of the open and distance learning system implemented, consists in a distributed repository** of *pedagogical documents* (or learning objects) of diverse granularity, origin, content, type, language, etc., authored by members of the participating Institutions with a variety of software tool (including those developed by the Project). The topology of the KPS is star-shaped, allowing for a simple mirroring scheme, performed every night across all nodes. The learning objects are indexed, usually by faculty staff, and stored in view of their use in telematics-based training or teaching curricula. The implied *share and reuse* strategy will work if (i) a critical mass of material is available at the time such curricula are composed and (ii) efficient search mechanisms allow for retrieving all (or at least most) relevant candidate resources in the KPS. The latter are obviously dependant on the indexation system which in turn must be designed according to the context of actual use.

1.1 ARIADNE: A Multicultural, Multilingual Application Context

Teaching curricula developed with this system are targeted at learners groups speaking at least six different languages. In

* ARIADNE (Alliance of Remote Instructional Authoring and Distribution Networks for Europe) is an RTD project, jointly funded by the European Union Commission and the Swiss Federal Office for Education and Science, in the 4th EU Framework Research Programme.

** The current implementation of the KPS relies on the ORACLE™ DBMS and comprises about ten nodes.

fact, several dozens educational applications have been developed and are currently under test in European Institutions and Organizations. While the critical mass of resources may take several years to be accumulated in the KPS, it is easy to foresee that large scale reuse will depend on teaching level arbitration, linguistic equivalency and *pedagogic and semantic similarity* seeking mechanisms. The teaching level arbitration issue has been partially addressed in Europe by the ECTS[4] Project, but not the basic problem of semantic and pedagogic interoperability which is crucial in the context of actual *teaching material* interchange.

1.2 Pedagogical Indexation

From the design stage of the KPS, the need for an adequate set of descriptors for the training resources under consideration became obvious. But could an existing library-originated indexation systems, be used to describe at least the *semantics* of these *electronic* teaching documents? After close inspection of the features of the two most promising thesauri: the Dewey classification and the Universal Decimal Classification (UDC), the idea was abandoned as (i) semantics appeared to depend on the fact that the resources were to be used for teaching; (ii) the vocabulary was far from being sufficiently detailed; (iii) the organization of knowledge domains seemed unusual and (iv) the corresponding thesauri of terms did not exist in ëcompatibleí versions in more than one language. This prompted us to design the ARIADNE *Pedagogic Header*, later recognized as a metadata set[***], to be used by indexers through the so-called *Pedagogic Header Generator* tool.

[***] We refer to the system implemented in the current (Sept. 1998) version of the ARIADNE tools, not to our ´†Educational Metadata Recommendation†ª [see: http://ariadne.unil.ch/metadata.htm] nor to the IEEE LTSC Metadata working document, based on ARIADNE-IMS work . [see: http://manta.readadp.com]

1.3 ARIADNE Educational Metadata

The current set of descriptors in the ARIADNE educational metadata is organized in **four categories**, two of which need little elaboration†: the so-called *general* and *technical* categories.

The ***general*** **category** regroups information about the origin, availability and metadata authoring of the resource. Most of the current descriptors of the Dublin Core can easily be mapped into it (with the notable exception of the *subject*). Interoperability will stem from the standardization of a conventional set of descriptors.

The ***technical*** **category**, absent from the Dublin Core and other classifications, is obviously needed when dealing with electronic resources that may run on a variety of platforms and operating systems, may or may not include multimedia components, may or may not exist in a compressed format and may have to travel long distances on networks of unspecified bandwidth. Thence the various descriptors that belong here. Interoperability is here at its usual, computer-related interpretation.

The two other categories address the core of our problem†: the so-called *semantic* and *pedagogical* characteristics. Leaving aside for the moment the issues related to multi-linguality, let us consider the descriptors that have been chosen and their implications on interoperabilities of various kinds.

2. Semantic Interoperability Mechanisms

The **semantic characterization** of a ´†learning object†ª relates to the knowledge (or the know-how) field in the context of which the teaching or training action *is to take place* (rather than *takes place*)†: we should mentally adopt here the attitude of the indexer who guesses at the kind of educational use that could be made of the resource under examination. Even if the indexer is not the document author, he must possess (and exert) considerable expertise to correctly classify it. One of the most prevalent and appealing ways of doing so is to adopt a hierarchically structured view of knowledge. We can then, first define†:

- the *discipline*, in the framework of which the teaching is to take place,

- then an optional sub-discipline,

- down to the *main concept* to be taught through the use of this resource.

As mentioned previously, no existing universal term classification could be identified from which items such as the *discipline*, and much less the *main concept* could be, under all circumstances, extracted. One can wonder why a systematic ontology such as the Dewey classification looks appropriate to librarians but unacceptable to pedagogues. The latter ideally would like to be able to enter descriptors coined in their own words†(this is especially true in Europe where traditionally teachers emphasize the personal colour they impart to their teaching). Adopting unrestrictedly such a way of defining semantic descriptors would inevitably lead to uncontrolled proliferation of often similar phrases in the conceptual metabase of the repository. This in turn would drastically reduce its usefulness.

One solution consists in defining a *fixed* list of disciplines after reconciling the needs of the Usersí Community. This is currently the case in our metadata tool where a fixed list of about a dozen ´†discipline pairs†[a] has been implemented[*].

Combating the proliferation of similar terms or phrases in the concept metabase implies some kind of guidance provided to the indexer. This guidance can only take the form of a facilitated reuse of terms *already* existing either in the concept metabase or in *ad hoc,* discipline-oriented, thesauri[5, 6]. Examples of such thesauri are the *Medical Subject Headings*[7] or the *ACM Computing Reviews Classification*[8], which are well accepted in the domain they address but not available in all fields of human knowledge or in many languages.

However, even when such thesauri exist, they are, after all, but *fixed* - however large - lists of terms, that cannot be expected to meet the requirements of the pedagogic indexers under all circumstances. These must then be allowed, when necessary, to enter their own phrase(s) as semantic descriptors. As a result, the concept metabase of the system will include both *free* entries and thesaurus entries. As our indexer population is not expected to be professionally trained in the art of ëidealí index-entry phrasing (i.e. will *not* systematically apply the usual rules concerning ordering of words, articles, plurals, etc.), some sort of supervision or coordination[9] will be needed to avoid the proliferation alluded to here above. Indeed, the ARIADNE KPS indexation mechanism implies a control step that is a prerequisite before the new metadata is released for replication across the repository, control to be performed by a domain expert, known as the *validator*.

Considered at the level of a single *local* node of the KPS, the role of the validator is relatively simple: at regular time intervals, he collects unvalidated headers pertaining to his expertise domain(s), verifies them and corrects, when appropriate, the semantic descriptors (as well as the other ones), either by putting them in a canonical form or by replacing them by close approximations already present in the *local copy* of the concept metabase. As this copy is refreshed every night across all nodes, the probability that *new* entries introduced at this particular local node on that particular day will duplicate similar *new* entries made elsewhere on the same day (and thus,

---

[*] This fixed list will evolve into a *restricted* list in the next version of our metadata tool.

unknown to the validator) is very small, albeit not nil. This small integrity risk is the price paid for dispensing from having a super-validation step performed, first at the central site of the KPS, to harmonize new inputs from all nodes, and then backward to these nodes to maintain full integrity. Note that this two step validation mechanism would call for three different states for the pedagogic headers: *unvalidated*, *locally validated* and *super-validated*.

Often, there are alternate ways of naming the main concept (the so-called *main concept synonyms*) and alternate important *other concepts,* dealt with in the pedagogical document. These are useful but optional information that can be entered by the indexer or even the validator.

Mechanism for bridging from any particular language to all others would call for a detailed discussion of the properties of *concepts*, considered as objects[10]. Let us mention here, briefly two important features:

(i) *disciplines* are both conceptual entries and attributes of other, ënormalí concepts; as they form a small set, multi-lingual translation of the discipline names is the obvious solution and can be done at system level;

(ii) *concepts* are considered as having *synonyms*, both in their particular language and in other languages; while entering a true synonym (in the same language) is *not* mandatory for the indexer, it *is* mandatory to enter at least one foreign language synonym (i.e. a translation), in one of a few *pivotal languages*. Thanks to the transitivity of the synonym relationship, the ´†sea of concepts†ª will gradually become criss-crossed by trans-lingual links. This will, in time, allow for retrieving all documents, whatever their language, dealing with any particular concept, whatever the language used for the semantic query.

Semantic interoperability here must be understood as the *correct identification* of all existing resources by a teacher querying the KPS by discipline or concept, i.e. agreement between this end-user and the indexers on the underlying teaching domain of the resources.

3. Pedagogic Interoperability Mechanisms

The **pedagogical characterization** addresses the need of describing as accurately as possible the target learner population (in terms of a teaching or training system) and a few basic and uncontroversial pedagogic parameters of the learning resource as well as its format.

Concerning the learnersí description, the only conclusion that one can draw from a survey of national training systems, is that there are approximately as many national conventional vocabularies for teaching levels as there are countries. On the other hand, there is general agreement (at least in western countries) on the broad context of the training. As a consequence, the following choices have been made :

(i) the indexer selects the *didactic context* from a list of a dozen terms (such as *secondary education, university undergraduate,..., continuous formation, professional training*, etc.); this list is easily translated at system level and provides for translingual bridging for this criterium;

(ii) a pragmatic, nationally oriented approach to the *level* is used: the indexer selects first the *country* he is referring to and then the *level* as expressed in *that* country for the selected didactic context; the bridging mechanism, built into the ARIADNE metadata tool will rely on a numerical equivalency scale relevant for each didactic context. The ECTS scheme, already mentioned, is an example of how this can be made to work.

The remaining pedagogic descriptors in our indexation scheme are: *type* <u>and</u> *format, semantic density* <u>or</u> *interaction quality* (depending on the type), and *pedagogic duration*. They appear to be rather uncontroversial and in any case easily translatable. Their interpretation is as follows:

The **type** is either :

expositive with possible values of the ***format*** taken from a list that includes items such as *text, hypertext, image, schema, video, sound*, etc. or

active with possible values of the ***format*** taken from a list that includes items such as *exercise, questionnaire, simulation*, etc.

The **semantic density** (for expositive material) or **interaction quality** (for active material) can take the values *low, medium and high*. Their meaning is self-explanatory.

The **pedagogic duration** is an estimate of the time needed by a typical learner to work through the resource under consideration.

Interoperability here can be thought of as *general agreement* on the *vocabularies* used for describing the former items, leading to good practices in the pedagogic uses of the learning resources.

4. Corporate Interoperability Issues

Extensions to the ARIADNE metadata scheme may be used to *accommodate **corporate training and information needs***. Semantic interoperability of corporate documents is becoming a crucial issue, because (i) enterprises need to disseminate, share and reuse corporate documents not only within and between their different departments but also with other organizations and (ii) on-the-job training is becoming vital for many companies. We found out empirically, by working with a number of different companies, that most corporate *training material* is built on the content of pre-existing corporate, *work-oriented documents*. This finding led us to consider attributing the status of ëlearning objectí to any corporate document that might acquire a pedagogic dimension. In order to accommodate corporate needs, extensions to the ARIADNE educational metadata scheme are however needed. They include corporate-oriented attributes such as *distribution list, aggregation level* and *confidentiality*.

The *distribution list* indicates the functions (rather than the names) of the people who are to receive the documents. To deal with ëproprietaryí denominations of functions across different (kinds of) corporations, one can then easily build equivalency tables to match different names given to a given function, as well as synonym tables that can take care of the translation in different languages.

The *aggregation level* corresponds to a hierarchical view of the document semantics. It relates to its more or less synthetical or analytical nature and usually depends upon the *type* of document concerned. Should there exist well-established document classification schemes in the relevant corporation, a simple mapping between these classifications could be performed.

The *confidentiality level* is partly derived from the class a document pertains to, but also from the kind of information it conveys. It might occur that information contained in a document should be regarded as *freely available* inside a company but otherwise as *highly confidential* outside the boundaries of the company. The final decision should generally be left to decide upon by the document author.

5. Current Implementation and Future Work

The ARIADNE system consists of five different *document-type-specific* authoring tools, *the pedagogical header generator* (or metadata tool), the *Knowledge Pool System* , the *Curriculum Editor*, the ARIADNE Management Interface (AMI) and the ARIADNE Learner Interface (AMI). Out of five, three authoring tools feature automatic *pre-header generation*, i.e. generate a large part of the metadata needed to index the resulting learning objects. The underlying rationale is that many data items are already known at authoring time and, especially in the case of serialized courseware production, ought not to be manually entered. The current version of the *pedagogical header generator*, a JAVA application, implements only part of the described mechanisms. In particular, support for auxiliary thesauri, translingual and other complex queries will only become available in the next version, due by spring 1999. However, upward compatibility will be preserved for the more than 1000 descriptions already produced. The *Knowledge Pool System*ís topology will evolve from the current simple star topology to a ësnow flakeí topology (i.e. a branched star), to better support scaling up in the number of users and nodes. The existing, simplified *Curriculum Editor* tool already relies on part of the metadata to generate the course images made available to learners through the ALI. The next version will also exploit them to support intelligently the process of curriculum creation and, incidentally, create the metadata proper to the curriculum description file itself. Finally the next version of the AMI will improve the management not only of the learners, but also of the database of indexers and course creators.

References

[1] E. N. Forte, M. H. K. Wentland Forte & E. Duval, *The ARIADNE Project, (part I) - Knowledge Pools for Computer Based & Telematics Supported Classical, Open & Education*, European Journal of Engineering Education Vol 22, No 1, pp. 61-74, March 1997.

[2] E. N. Forte, M. H. K. Wentland Forte & E. Duval, *The ARIADNE Project, (part II) - Knowledge Pools for Computer Based & Telematics Supported Classical, Open & Education*, European Journal of Engineering Education Vol 22, No 2, pp. 153-166, June 1997.

[3] K. Cardinaels, K. Hendrickx, E. Vervaet, E. Duval, H. OliviÈ, F. Haenni, K. Warkentyne, M. Wentland Forte & E. Forte, *A Knowledge Pool System of Reusable Pedagogical Elements*, in C. AlvegÂrd (ed.), *Proceedings of CALISCEí98*, pp. 54-62, 4th International Conf. on Computer aided Learning and Instruction in Science and Engineering , Gˆteborg, Sweden, 1998.

[4] ECTS Users Guide, see: http://europa.eu.int/en/comm/dg22/socrates/guide-en.doc

[5] E. Mena, V. Kashyap, A. Illarramendi, and A. Sheth, *Domain Specific Ontologies for Semantic Information Brokering on the Global Information Infrastructure*, Intl. Conf. on Formal Ontology in Information Systems, (FOIS'98), Treno, Italy, 1998.

[6] V. Kashyap and A. Sheth, *Semantic Heterogeneity in Global Information Systems: The Role of Metadata, Context and Ontologies*, in M. Papazoglou, and G. Schlageter, (Eds.), *Cooperative Information Systems : Current Trends and Directions*, Academic Press, 1997.

[7] Medical Subject Headings, see: *http://www.nlm.nih.gov/mesh/meshhome.html*

[8] ACM Computing Classification, see: *http://www.acm.org/class/1998/*

[9] A. Ouksel & C. Naiman, *Coordinating Context Building in Heterogeneous Information Systems*, Journal of Intelligent Information Systems vol. 3 No 1, pp. 151-183, 1994.

[10] E. Forte & F. Haenni, *Structure de la base de concepts dans ARIADNE*, internal report (in French), ref. ARIADNE/EPFL/WP16.101/1.0, EPFL-LEAO, 1998