Erik Duval, Eddy Forte, Kris Cardinaels, Bart Verhoeven, Rafael Van Durm, Koen Hendrikx, Maria Wentland Forte, Norbert Ebel, Maciej Macowicz, Ken Warkentyne, and Florence Haenni

THE ARIADNE KNOWLEDGE POOL SYSTEM

This European-based distributed digital library of educational resources delivers reusable components to hundreds of multicultural/multilingual teachers and learners.

With the exploding popularity of the Internet and Web, opportunities for useful educational applications of digital resources have also increased dramatically. However, developing and exploiting these applications entail many challenges. Producing quality content is difficult and costly. Whereas academic researchers are used to producing written material, producing rich digital learning resources may require a multidisciplinary team with a background in pedagogy, graphical design, computer programming, and other disciplines. Moreover, the integration of digital learning support with more traditional paper-based material and face-to-face lectures in a pedagogically sound way is not straightforward.

Digital libraries of educational resources can help overcome some of these challenges [8], saving time and energy while avoiding the continuous rebuilding from scratch of similar materials. Such a library would at least make it easier to learn from what others have already experienced. Educational digital libraries could also make the practical management and social appraisal of learning materials more like the management and appraisal of research material. In particular, creating quality learning resources would be regarded (and possibly rewarded) like scientific publications. An example of an initiative along these lines is the recently introduced *ACM Journal of Educational Resources in Computing*, which encourages the submission of learning materials by treating them as journal submissions.

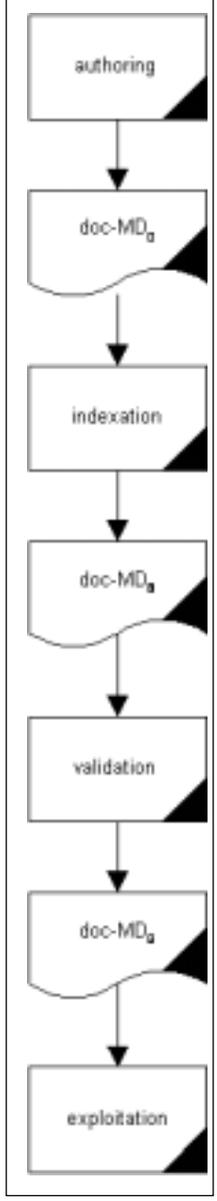
In order to promote the concept of "share and reuse" for educational resources, ARIADNE, the European educational digital library project, was initiated in 1996 by the European Commission's telematics for education and training program. Since then, an infrastructure has been developed in Belgium and Switzerland for the production of reusable learning content, including its description, distributed storage, and discovery, as well as its exploitation in structured courses [6, 7]. The core of this infrastructure is a distributed library of digital, reusable educational components called the Knowledge Pool System (KPS) now actively used in both academic and corporate contexts.

KPS content can include "anything digital" without format restriction. When a conceptual document consists of more than one file (such as an

HTML document with embedded images), all relevant files are included in one compressed (zip) file. The ARIADNE community distinguishes between active documents requiring a reasoned action by the learner (such as a questionnaire or an exercise) and expositive documents (such as a text or video clip) requiring the learner to read, listen, or watch. Active documents add great value to conventional learning; they are also the most difficult to produce, and hence the most valuable for reuse.

The KPS includes descriptions (metadata), as well as the documents themselves, making it easier to replicate documents across all nodes of the system while ensuring convenient access without excessive download times. It also means the KPS can act as a reference library from which a copy of a document can always be retrieved. In 2000, the KPS contained about 2,500 documents; roughly 25% of them were active, most in the form of exercises, which are relatively easy to create. "Really active" documents, like multiple-choice questionnaires, quizzes, auto-evaluations, and simulations, account for about 10% of overall KPS content. Most of the expositive documents were text (40%), hypertext (30%), slide sets (15%), and video clips (11%). With respect to document formats, Microsoft Word and HTML each represent about 25% of KPS documents. Formats for the other 50% include images (GIF and JPEG), video (MPEG and QuickTime), and applications (PowerPoint and Toolbook).

Figure 1. A document's life cycle and metadata.



Knowledge Pool Community

In ARIADNE, which takes its name from the mythological Greek figure who left a thread to help find his way in a maze, maintaining KPS quality is mostly a community issue supported by way of different roles for users. These roles reflect a document's life cycle and associated metadata (see Figure 1); individual users can assume more than one role.

The "author" is concerned with the production of his or her documents. Some authoring tools automatically generate some of the metadata (indicated as MD_0 in the figure), including date of production, file format, and size.

An "indexer" describes all relevant aspects of the document using an indexation tool (see Figure 2). An indexer is often involved in the authoring phase as well; the content-production process is the most appropriate time to capture metadata. When external or existing material is reused, indexation becomes an independent process in itself.

A "validator" downloads the document from the KPS, verifying that all relevant files (such as images embedded in HTML files) are included and that applications run in the hardware and software environment identified in their metadata. The validator does not verify the document's content or pedagogical style; it is practically impossible to achieve consensus in a cross-domain community like ARIADNE's on quality criteria for content and pedagogical approach.

Originally, during the experimental phase of ARIADNE development, documents were exploitable only when they passed the validation stage. However, this quality check created a bottleneck in KPS operations, as there were only a limited number of validators responsible for reviewing documents in their domains of expertise. The ARIADNE community eventually decided to make available all documents, indicating in the user interface whether or not the ones included in search results are validated, so end users can assess the trustworthiness of the metadata.

A "course manager" exploits KPS content, typically involving the specially developed ARIADNE query tool (see Figure 3) to identify relevant documents while integrating the documents into a course. In ARIADNE, we developed a course editor for this purpose; the teacher automatically generates a course Web site, including the documents in a meaningful pedagogical context [9]. "Students" normally use the course Web site to access the documents, though they could also use the KPS as a background library when undertaking self-directed discovery.

The "general public" has access only to the metadata, because the ARIADNE community wants to maintain a level of trust, helping authors feel at ease about contributing their content to the KPS for free. About 88% of ARIADNE documents are freely available. The main usage restriction is that commercial exploitation of free documents without prior consent from the rights holder is not allowed. The intellectual property rights remain where they resided before the document was entered into the KPS. For most ARIADNE content developers, ensuring that only identified users can access documents is an important guarantee that their rights are being respected.

Figure 2. The ARIADNE indexation tool.

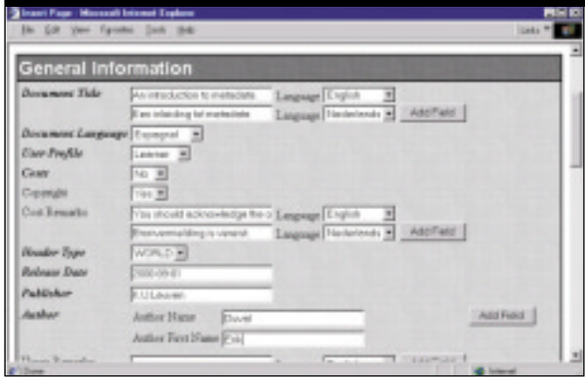


Figure 3. The ARIADNE query tool.

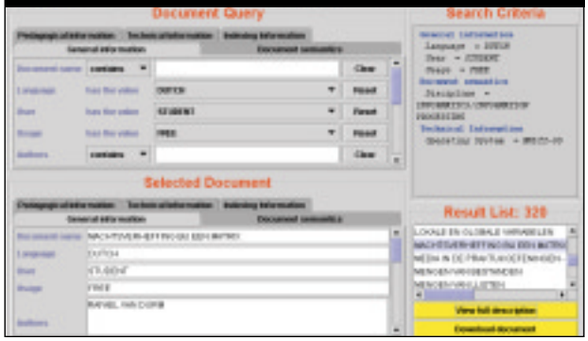
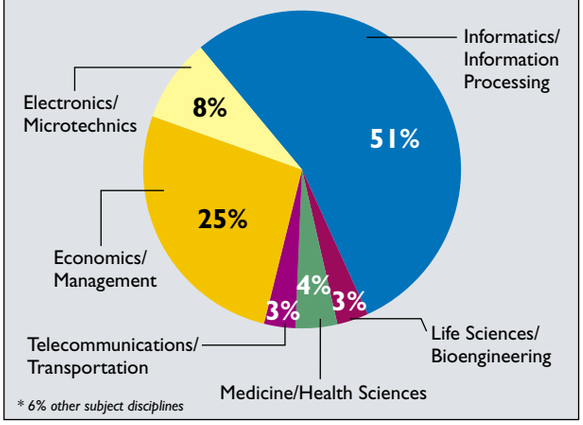


Figure 4. Subject disciplines in the Knowledge Pool System.



Metadata

Metadata is extremely important for the proper management of any resources [3]. It should be relatively easy to generate metadata, either manually or semiautomatically, but metadata should also make it easy for an end user to identify relevant material.

The typical end user of the ARIADNE KPS is a teacher using educational documents in a course. We are thus more focused on structured formal learning, as exemplified by an academic context [10]. For such formal learning processes, we feel the ARIADNE Learning Management System should not completely eliminate the tutor's or teacher's interventions, however, as research into adaptive educational systems relying more on automatic matching of learners with content is still in its early stages. Nevertheless, there are circumstances in which it is appropriate to give access to the KPS to autonomous students, just as it can be appropriate to refer them to conventional libraries.

As a library, the KPS requires traditional metadata, including title, authors, and publication date. In order to describe the semantics of the document, the ARIADNE community has developed its own simple hierarchical scheme of disciplines and subdisciplines; existing classifications are not well developed for some domains and not available in all relevant languages, though developing such classifications or alternative mechanisms is an important area of future research. KPS content today is oriented more toward technical sciences, strongly represented by computer science (50%), economics (25%), and electronics (8%) (see Figure 4). This breakdown reflects the original composition of the ARIADNE consortium, though it is gradually reflecting a more even representation. Moreover, all documents are assigned to a particular subdiscipline, and their main concept, synonyms, and, if applicable, secondary concepts are defined.

Due to its digital nature, the KPS includes additional metadata describing the technical characteristics of documents, including a unique identifier automatically generated by the system, information about MIME types in the document, the uncompressed size of the document, and requirements with respect to the computing platforms (so end users can assess whether learners are able to view the particular document). So-called "meta-metadata" indicates when documents were inserted into the KPS, who described them, and whether or not they were validated.

Due to the KPS's distributed nature and European context, support for multiple languages is extremely important [7]. English and French each account for more than 30% of the documents; Italian, German, and Dutch each account for about 10% of the documents. Because these statistics still reflect the composition of the original ARIADNE consortium, they are not completely representative of the European context. However, they do illustrate that the content of the KPS is not dominated by one human language. The same requirement applies to the metadata; in fact, values for each of the free-text metadata fields can be entered in more than one language, as in Figure 2. In order to support rights management, the metadata also indicates whether any costs are associated with the use of a particular document, along with any applicable usage restrictions.

Figure 5. Geographical deployment of the Knowledge Pool System.

The most innovative original ARIADNE work is the educational metadata [3, 7], including:

- Document type (active or expositive);
- Format (questionnaire, simulation, hypertext, and others);
- Usage remarks (explaining how the document can be used in a pedagogically sound way);
- Didactical context and course level (describing the kind of learners for whom the document is intended);
- Difficulty level, interactivity level, and semantic density (qualitative indicators range from "very low" to "very high"); and
- Pedagogical duration (time typically needed by learners to work with the document).

Metadata is inherently subjective; a document that is easy to use for some indexers may be rather difficult for others indexers. This subjective nature applies not only to the educational category; for example, all indexers won't always agree as to what constitutes a document's main concept.

Client and Server Tools

End users interact with the KPS through client tools; Java and Web applications insert documents and their associated metadata into the KPS, search for relevant documents, and download them from the KPS. The Web application in Figure 2 was developed in Java Server Pages (JSP) to interact with the KPS through the Structured Query Language (SQL). Special care during development of these tools ensured they would be adaptable; changes to the underlying structure of the KPS and to the user interface (including translations) are relatively easy to accommodate.

Java applications interact with the KPS through either an XML stream or JDBC/SQL. With XML, they interact with KPService, a middleware application we developed, also in Java. Figure 3 illustrates the ARIADNE query tool, which identifies and downloads relevant documents. More advanced features can be supported in Java applications, as opposed to Web forms. However, installing applications means extra complications and effort for many of our users, especially where Java runtime environments have to be installed and configured.

The ARIADNE server architecture is based on an (Oracle) relational database management system that takes care of metadata management and basic query processing. The actual documents are stored as files on a Web server (Apache) providing user authentication for protected documents.

We adopted a tightly coupled architecture in which metadata and documents are all stored in a central knowledge pool (CKP), rather than in a heterogeneous set of loosely coupled search engines. Thus, we do not have to support federated searches, which would be sent to more than one server. Due to bandwidth limitations in the connections to the CKP, and, more generally, in order to make local sites more self-supporting, we replicate all metadata (but only for locally relevant documents) at all sites. Although our approach may seem less scalable than federated systems, it will be a long time before the number of educational resources exceeds the hundreds of thousands the current architecture is designed to cope with.

Replication of resources is based on a three-level star-shape topology, in which so-called regional knowledge pools replicate from and to the central knowledge pool. End users interact with local knowledge pools that in turn interact with regional ones. Figure 5 illustrates the current structure of the KPS, with central knowledge pool (in red), regional knowledge pool (in green), and approximately 20 local installations (in blue) throughout Europe. Extensions to Latin America and North Africa are being discussed.

If we want the KPS to be a valuable repository, it has to reach a critical mass of users. In addition, the interchange of metadata with other repositories is highly desirable, thus, our continued involvement in the IEEE Learning Technology Standardization Committee (LTSC) Learning Object Metadata (LOM) Workgroup (see ltsc.ieee.org/wg12/index.html). The LOM standard is based on earlier ARIADNE work, as captured in a submission made to the LOM workgroup in collaboration with the Instructional Management Systems Consortium [2, 4]. We are also active in localization and internationalization issues, both in the

IEEE/LTSC and the European Committee for Standardization Learning Technology Workshop (see www.cenorm.be/isss/Workshop/lt/Default.htm). Our goal is to foster interoperability between learning platforms while protecting humanistic values, including multiculturalism and learner privacy.

Although ARIADNE started in 1996 as a European project, it is now a not-for-profit, international foundation that collaboratively maintains and develops its metadata and learning tools. An open-source approach toward tool development is being investigated.

Related Work

The Networked Digital Library of Theses and Dissertations (see www.theses.org) is more limited than ARIADNE in scope, as it is restricted to theses and dissertations from around the world. Architecturally, it operates on a rather different model from the KPS of loosely coupled sites, so searches need to be federated or based on harvesting, as in the Open Archives Initiative (see www.openarchives.org). The Computer Science Teaching Center (see www.cstc.org) is more similar to the ARIADNE KPS, as it attempts to build a digital library of computer science teaching resources. Although its review process is more elaborate, its metadata set is much simpler, and the number of documents in CSTC today is limited.

The U.S.-based Science, Mathematics, Engineering, and Technology Education initiative is also building a digital library (see www.smete.org). Registered users can add comments on resources, a feature supported in ARIADNE, but that has attracted limited user activity. It should be relatively straightforward to establish interoperability between ARIADNE and SMETE, as the latter generates LOM instances in XML. From the KPS, ARIADNE can do the same, as well as import metadata records.

The Education Network Australia (EdNA) supports a repository of 10,000 evaluated resources called EdNA Online (see www.edna.edu.au). Its metadata structure is based on the 15 Dublin Core metadata elements, extended with nine EdNA-specific elements. These nine elements relate mainly to meta-metadata and reviews of resources. One pedagogical metadata element—the "user level"—is included, and searching is supported over only five fields. EdNA metadata is typically stored in HTML metatags, and items may be suggested by submitting their URLs.

The metadata structure of the Gateway to Educational Materials (GEM) digital library (see www.geminfo.org) is also based on the Dublin Core metadata element set but includes more pedagogical attributes adopted from the IEEE LTSC LOM work. As in EdNA, only a limited number of elements, in this case five, are searchable. The organizational structure is similar to ARIADNE's, requiring consortium members to not only consume but contribute to the resources accessible through the gateway. Neither EdNA nor GEM include the actual document in the server.

Open Issues

Although the KPS is an operational system supporting the sharing and reuse of educational resources, a number of open issues remain, some of which are subjects of our current research. Also worth mentioning is that these issues are relevant in the context of software reuse [1, 11].

For example, a puzzling issue involves appropriate document granularity; aggregates and their constituent parts can be described separately. Smaller documents are indeed more reusable, because they incorporate less pedagogical context. However, the value end users derive from reuse increases, as the document involves more elaborate content.

A difficult issue involves the customizability of resources, including language translation and, important in technical fields, the substitution of mathematical symbols so they are consistent throughout a course. This apparently simple operation, which can be supported by "resource bundles" and configuration files, can pose user interface problems, as, say, text becomes longer when translated. If the content needs to be customized as well, then it is necessary to import the document in the original or at least in a compatible authoring environment. As standards for content (such as XML-based document formats) are more widely adopted, translating content could become more feasible.

A more intricate issue is that of design for reuse; documents are more readily reusable if they are self-contained and can be integrated in new learning contexts. In practice, what this implies is not always clear, though some guidelines can be formulated; for instance, an author should not link directly to other documents in a course, as these documents may not be available in the context of reuse.

Finally, there are limitations as to the kind of resources ARIADNE supports easily. Tight coupling between documents, as when a simulation model is linked with independently developed graphical representations, requires more sophisticated document models (such as JavaBeans) than the ARIADNE black-box approach [12]. However, restricting ARIADNE authors to such document models would limit ARIADNE's scope. Moreover, such a selection would have to be reassessed continuously in light of the frantic pace of technological evolution.

ARIADNE relies mostly on a download (rather than a streaming) paradigm for document delivery, though

some work by ARIADNE developers has been done on integrating a video-streaming server as an extension to the KPS. This is only important if one wants to include, say, full retransmissions of lectures in a distance course.

Describing educational documents is tedious. As mentioned earlier, metadata can be gathered semiautomatically during content production, especially when using document-type-specific authoring tools. Even in such cases, semantic and pedagogical metadata is often difficult to generate automatically and needs to be verified through human intervention. When new documents are based on existing documents that already include metadata, templates based on the existing documents can be reused to describe the new document.

The main mechanisms for semantic interoperability include thesauri and vocabularies for metadata elements [7]. In a multicultural environment, where educational context, policy, and organization are likely to diverge, users should be able to search for appropriate learning resources in their own language, using their own terms of reference for educational characteristics. That is why, in the context of the European Committee for Standardization Learning Technologies Workshop, we are researching relevant taxonomies and vocabularies, as well as their applicability and interrelationships. One aim is to establish mappings between the taxonomies of different communities.

Conclusion

We do not suggest that distributing documents to learners is the only requirement for education. The ARIADNE toolset enables those responsible for courses to plan interactions between teachers and learners, or among learners. Nevertheless, a digital library of reusable educational components enables the sharing and reuse of such components, which remain costly to produce. In order to reach the required critical mass of teachers, learners, and reusable components, a library has to be widely deployed. Support for the community of diverse users is another requirement. For these and other reasons, the KPS enables its users to integrate technology-based components in the educational practice in most lingual and cultural contexts. C

References

1. Batory, D. and Poulin, J., Eds. *Proceedings of the 9th International Workshop on Institutionalizing Software Reuse* (Austin, TX, Jan. 7–9, 1999); see www.umcs.maine.edu/~ftp/.wisr/wisr.html.
2. Duval, E. An open infrastructure for learning (the ARIADNE project): Share and reuse without boundaries. In *Proceedings of Enabling Network-Based Learning* (E. Rämö, Ed.) (Helsinki, Finland, June 2–5). Espoo-Vanta Institute of Technology, Espoo, Finland, 1999, 144–151; see www.enable.evitech.fi/enable99/papers/duval/duval.html.
3. Duval, E., Vervaet, E., Verhoeven, B., Hendrikx, K., Cardinaels, K., Olivié, H., Forte, E., Haenni, F., Warkentyne, K., Wentland-Forte, M., and Simillion, F. Managing digital educational resources with the ARIADNE metadata system. In *Metadata and Organizing Educational Resources on the Internet*, J. Greenberg, Ed. Haworth Information Press, New York, 2000, 145–171; see also *J. Internet Catalog. 3,* 2/3 (2000).
4. Haenni, F., Warkentyne, K., and Forte, E. *ARIADNE Pedagogical Header: Structure & Description*, ARIADNE Proj. Rep. WP3.101/1.3, June 1996.
5. Forte, E., Wentland-Forte, M., and Duval, E. The ARIADNE project (Part 1): Knowledge pools for computer-based and telematics-supported classical, open, and distance education. *Europ. J. Eng. Edu. 22,* 1 (1997), 61–74.
6. Forte, E., Wentland-Forte, M., and Duval, E. The ARIADNE project (Part 2): Knowledge pools for computer-based and telematics-supported classical, open, and distance education. *Europ. J. Eng. Edu. 22,* 2 (1997), 153–166.
7. Forte, E., Haenni, T., Warkentyne, K., Duval, E., Cardinaels, K., Vervaet, E., Hendrikx, K., Wentland-Forte, M., and Simillion, F. Semantic and pedagogic interoperability mechanisms in the ARIADNE educational repository. *ACM SIGMOD Rec. 28,* 1 (Mar. 1999), 20–25; see www.acm.org/sigmod/record/issues/9903/ index.html.
8. Fox, E. and Marchionini, G. Toward a worldwide digital library. *Commun. ACM 41,* 4 (Apr. 1998), 28–32.
9. Macowicz, M., Warkentyne, K., and Forte, E. The ARIADNE Web-based learning environment: An overview. In *Proceedings of Web-Based Learning Environments 2000* (Porto, Lisbon, Portugal, June 5–6). University of Porto, Portugal, 2000; see www.fe.up.pt/merlin2000/ wble2000/.
10. Marchionini, G. and Maurer, H. The roles of digital libraries in teaching and learning. *Commun. ACM 38,* 4 (Apr. 1995), 67–75.
11. Prieto-Diaz, R. Implementing faceted classification for software reuse. *Commun. ACM 34,* 5 (May 1991), 88–97.
12. Roschelle, J., DiGiano, D., Koutlis, M., Repenning, A., Phillips, J., Jackiw, N., and Suthers, D. Developing educational software components. *IEEE Comput. 32,* 9 (Sept. 1999), 50–58.

Erik Duval (erik.duval@cs.kuleuven.ac.be) is a professor in the Departement Computerwetenschappen, Katholieke Universiteit/ Leuven, Belgium.

Eddy Forte (eddy.forte@epfl.ch) is chair of the ARIADNE Foundation and heads the Laboratoire d'Enseignement Assistée par Ordinateur, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Kris Cardinaels (kris.cardinaels@cs.kuleuven.ac.be) is a Ph.D. candidate in the Departement Computerwetenschappen, Katholieke Universiteit/Leuven, Belgium.

Bart Verhoeven (bart.verhoeven@cs.kuleuven.ac.be) is a Ph.D. candidate in the Departement Computerwetenschappen, Katholieke Universiteit/Leuven, Belgium.

Rafael Van Durm (rafael.vandurm@cs.kuleuven.ac.be) is a Ph.D. candidate in the Departement Computerwetenschappen, Katholieke Universiteit/Leuven, Belgium.

Koen Hendrikx (koen.hendrikx@cs.cs.kuleuven.ac.be) is a Ph.D. candidate in the Departement Computerwetenschappen, Katholieke Universiteit/Leuven, Belgium.

Maria Wentland Forte (mwentlan@hec.unil.ch) is a professor in the Knowledge Management Unit, Université de Lausanne, Switzerland.

Norbert Ebel (norbert.ebel@epfl.ch) is a postdoctoral researcher in the Laboratoire d'Enseignement Assistée par Ordinateur, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Maciej Macowicz (maciej.macowicz@epfl.ch) is a postdoctoral researcher in the Laboratoire d'Enseignement Assistée par Ordinateur, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Ken Warkentyne was a postdoctoral researcher in the Laboratoire d'Enseignement Assistée par Ordinateur, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

Florence Haenni was a Ph.D. candidate in the Laboratoire d'Enseignement Assistée par Ordinateur, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

We gratefully acknowledge the financial support of the European Commission Telematics for Education and Training Program and of the Swiss Federal Office for Education and Science.