Towards Integration of Learning Objects Metadata and Learner Profiles Design: Lessons Learnt from GESTALT

M. Konstantopoulos, J.S. Darzentas, P. Koutsabasis, T. Spyrou,
and J. Darzentas
Department of Information and Communication Systems, University of the Aegean,
Samos, Greece

ABSTRACT

Within the context of learning technology environments, there is much effort on one hand, at defining metadata for educational content, and on the other, at specifying learner user profiles. However, less attention has been paid to the understanding of the relationship between these two areas of research, in terms of semantic and structural correlations. Based on an implementation of learning object metadata and learner profiles in a learning technology system undertaken in the Getting Education Systems Talking Across Leading Edge Technologies (GESTALT) project, this paper argues that the definition of a metadata schema and a User Profile should be a joint effort due to the complementary nature of these two descriptions in regard to the critical functions of the educational process.

INTRODUCTION

The transformation of education from location based to distance on-line based has dramatically increased the interest and research in the area of distance learning and education. A useful definition (CDLP, 1999) states that: "Distance Learning is an instructional delivery system which connects learners with educational resources". This definition emphasises two main entities in the learning process: the learners and the learning materials. One may argue that instruction is also part of the process, however, there is no doubt that one of the main tasks of any educational process is to provide the learner with suitable learning materials.

Correspondence: M. Konstantopoulos, Department of Information and Communication Systems, University of the Aegean, GR 83200 Karlovassi, Samos, Greece.
E-mail: {mpk, jennyd, kgp, tsp, idarz}@aegean.gr

This task is accomplished in on-line systems using some type of platform of educational technology, variously known as computer-based training, electronic performance support systems, computer-assisted instruction, intelligent tutoring, education and training technology (Farance, 1999). Although the emphasis may vary according to the specific approach, the learner and the educational resources remain principal parts of the equation.

Recent trends in the pre-standardisation activities regarding Learning Technologies specify, among other things, special types of representation for these main entities. Learning materials are described by structured formats, known as metadata, while learners are characterised by learner or user profiles. In order for on-line learning materials to be delivered to a user within an on-line learning environment, some form of interoperation of the metadata and the profile needs to take place.

In most current systems, the correlation between the User Profile and the learning material descriptions is not well developed, if at all. Most often, the system operates by supplying learners with planned programmes of material, taking into account only basic user characteristics. However, as learners come to on-line learning environments to take modular courses rather complete programmes of study, especially in distance learning programs that are addressing vocational and life-long training needs, the requirement for customisation of the material to the learner becomes more imperative.

The need for well-designed profiles, and adequate metadata descriptions of content and their combined use to enable the delivery of appropriate learning content to users, is the subject of this paper. It describes the lessons learnt from the implementation of a learning technology system, within the context of an ACTS project Getting Education Systems Talking Across Leading Edge Technologies (GESTALT), where, as well as work on extensions and modifications to learning object metadata, User Profiles and their interaction with the metadata, were proposed.

More specifically, the GESTALT project designed and implemented an extended open on-line educational environment by creating an integrated trial system, which encompassed the whole of the delivery of distributed on-line learning from searching for learning courses and resources, enrolling on and following programs of study delivered over networks, and providing mechanisms for passing information between various parts of the system in a secure and trusted manner. With regard to metadata and User Profiles, the project implemented some of the most up to date specifications. Extended versions of the learning object metadata (LOM) and the public and private information

(PAPI) specifications were used to describe the learning materials and the learner profiles, respectively. The GESTALT trial system based its retrieval mechanism on these two descriptions and detected possible design drawbacks in these specifications in terms of interaction between the User Profiles, and the Metadata.

This paper argues that while the work on specifications is producing very useful and promising results, these two strands of work should be more closely integrated in order to maximise the benefits of retrieval functionalities. The next section gives an overview of content metadata concepts generally, as well as with regard to learning technology, while the Section discusses current 'USER PROFILES IN EDUCATION' issues and how these are being reflected in learner profiles. In Section 'CURRENT TRENDS IN METADATA AND PROFILES IN LEARING TECHNOLOGY TECHNOLOGY SYSTEMS' these technologies are positioned in an abstract model of a learning technology system (the LTSA), in order to understand the information stores and flows. The Section 'THE GESTALT CASE' discusses the GESTALT case, and its implementation and integration of available specifications on metadata and User Profiles focusing on their inter-operation in the context of a Resource Discovery Service (RDS). Finally, 'CONCLUSION AND FUTURE WORK' are given in the last Section.

CONTENT METADATA IN EDUCATION

The use of metadata in information theory is not a new concept. Although, not exactly similar in their purposes and scope, information classification and information indexing can be seen as a first attempt to keep 'data on data'. Categorizing or describing the available information in order to efficiently disseminate or retrieve it has been an issue since Aristotle and his efforts at the library of Alexandria and it concerns every scrap of human knowledge publicly exposed in any shape or form. Knowledge in digital form in combination with technological advances permitting huge collections of knowledge and wide circulation of every piece of it, have pushed the issue of metadata to the fore and a subject of concern not only to librarians, the traditional guardians of knowledge, but also to educationalists (Darzentas, 1999) and those concerned with data storage and retrieval (Grotschel & Lugger, 1998).

The most common definition of metadata is that 'Metadata is data about data'. However, in a learning object context, since the metadata's main purpose is to relieve the potential users of the data objects from having to have full advance knowledge of an object's existence or characteristics (Dempsey & Heery, 1997), the metadata describe structures and functions like location, discovery, documentation, evaluation, selection, etc. Metadata is required to be 'machine processable' information. Its structure is its main usability advantage in the way that the data represented by metadata can be collected, processed, analysed based on the metadata's well-defined semantics and structure (Berners-Lee, 1997).

Metadata is then another tool in the effort to manage the bulks of resources available on the net. According to the Instructional Management Systems project (IMS, 1999:2): "A metadata specification makes the process of finding and using a resource more efficient by providing a structure of defined elements that describe, or catalog, the learning resource, along with require-ments about how the elements are to be used and represented''. However, it might be argued that the existing technologies for managing data, in general the two broad terms 'indexing' and 'classification', behave adequately without basing their efficiency on metadata concepts. If a comparison between the work done on classification from the previous century and the Dewey Decimal Classification Scheme (Koch, 1997) until now and the Yahoo! (Steinberg, 1996) Web site is made, it can be seen that there is a well-established area of research. Moreover indexing has been used widely in the database industry and web sites, and the research is moving constantly to more efficient solutions (Lawrence et al., 1999).

However, the reasons for the metadata trend can be based on some generic technology-related factors such as: (a) the need to increase automation of the production of descriptions, (b) the need to reduce the semantic divergence of these descriptions in a complex environment, and (c) the need for common descriptions whose access is device independent. The first two claims are based on the fact that in this networking age, the content to be described is likely to be located at a variety of locations. While once the content to be described was located in pre-defined places (e.g. the libraries) where experts (e.g. librarians) performed the extraction or manipulation of the descriptions in an appropriate manner, nowadays with the diversity in the nature of the content, as well as the widening of the traditional roles of content providers, the need is greater than ever for some consensus on metadata schemes as well as and metadata production (or extraction) tools. With regard to the need for

device independence, from the client (user) side there is a tendency towards the production of new languages (W3C, 2000) that will permit the access of information from a plethora of devices.

Given such a distributed environment with regard to the production of data (content) and metadata is not to say that the era of common relational databases is over, as metadata remains data. What it does mean is that the administrative load for the management of the resources can be significantly reduced by the fact that an object of information (i.e. a resource) can move around the world from the first moment without first having to be recorded as database information. The description of the information has already been produced by the author using a well-defined schema. Still the semantics of this schema (i.e. the metadata schema) remains an open question. While name-spaces (the most common example is the XML namespaces) resolve issues of ambiguity and conflicts, they do not clearly define the semantics of the information. This can only be done when using well-established schemes with obvious semantics. The Dublin Core (DC) is perhaps the best known current work using an easily understood scheme. Given that its complexity does not exceed that of a basic library catalog card, most of its elements have, by analogy, a commonly understood semantics. But although the semantics of such a generic and evident schema are clear, the drawback with such generic schemas is the lack of information specific to the area of metadata exploita-tion. A solution to this problem has been to create specific schemas created for specific context.

These specific schemas demonstrate the real benefit of metadata. Position-ing metadata in a specific application field — in this case, that of distance-learning — adds significant value both to the structure and the application because metadata can become a powerful tool in the structured representation of the functionality of a system. For example, the IEEE LTSC (2000:1) perceives some of the aims of developing learning object metadata to be a way ''to enable sharing and exchanging of learning objects across any technology-supported learning system'', ''to enable computer agents to automatically and dynamically compose personalized lessons for an individual learner'' and ''to complement the direct work on standards that are focused on enabling multiple Learning Objects to work together within an open, distributed, learning environment''. Against this background of a conceptual framework of future applications, the IEEE has developed a metadata schema capable of reflecting and supporting the functionality of the framework. The work described here subscribes to this perception that the learning object metadata

is expected to support a machine functionality that is going to implement personalized educational programs based on a description of the learner. In this description, the learner profile is thus essential to the exploitation of metadata of educational content.

USER PROFILES IN EDUCATION

The User Profiles are basically the 'data on a person'. In a digital environment, where there is a need to deliver user customised environments and services, primarily referring to user specific (i.e. adaptable to and by the user) interfaces and processes, there is a need to keep user records. Records are usually static structures containing only an instance of the user related information basically of the moment that this information was first collected or last updated. Depending on the implementation and the intended use of the profile, this scenario may differ in the fact that a record may contain previous (outdated) states (information) of the object or that some data on produced 'on the fly' (e.g. user machine's data) and have very limited lifetime. The adaptivity of the profile is closely linked to the function, such a profile is supposed to accomplish the environment this profile is related to. For example, in the GESTALT project, an initial registration phase of the learner was followed by automatic updates of the profile by the Learning Environment when specific milestones on learner's progress had been attained.

Outside of education applications, the User Profiles are currently contained as a functional parameter in most multi-user operating systems, applications and distributed environments. A common way to implement User Profiles is by using a directory service. Directory service is at present enjoying a renaissance of interest, and thus is another 'redefined' term in the technological literature. Whereas directory services once meant 'white pages', currently the term tends to substitute or amalgamate with the traditional user management of an operating system (SUN, 2000). So while once directories were used to store data that were to be presented in a catalog (human and material resources of an organization mainly), now directories are unified with the operating systems, they hold operating system User Profiles, they define the access rights of the information system entities, etc. Technology is now unifying operating systems, mail servers and web server user records and the trend is to expand these directories in order to store profiles for as many applications as possible running in a digitised system (BED Web Team, 1999; SUN[TM], 2000).

In the area of Learning technology the term most commonly used is 'learner profiles'. Efforts at defining a common learner profile scheme to be exchangeable across the different learning systems have been undertaken by IEEE. Farance and Schoening (1998) introduced the PAPI specification which is a learner profile and a data interchange specification, for describing 'portable' learner (student) records. An implementation of PAPI has been undertaken in the GESTALT project (Konstantopoulos et al., in press) and is discussed in the Section 'THE GESTALT CASE'. Another issue regarding the data on a learner is the management of student identifiers (IEEE LTSC, 2000:2). One strand of the IEEE work is focusing on 'enabling schools and other non-technical organizations to easily issue unique student identifiers to their students or members' and 'enabling schools and vendors to offer learning services through distributed management systems and distributed student databases'.

While these efforts are one step towards the standardization of the learner information in the learning systems, they do not tackle an issue that will be critical if learning systems are to be exploited by educational organizations, namely to find ways to integrate the newly introduced learning systems with the existing organizational infrastructure. One scenario would be to have just one student record in a database, which would provide to the student the access to all institutional facilities and would be usable by all the institutional management tools, such as the computer labs, the institutional student records, the on-line learning environments and whatever else will come up in the future. Such a scenario would dramatically reduce the administrative load in an institution. One central database would be accessed by any application to provide the student all the access rights needed. Such a scenario cannot be implemented without the use of a directory service (Microsoft, 2000).

The speed with which technology is moving points to a situation where student records would be some kind of operating system user records (Microsoft, 1999). This is especially so, since the directory services protocol and corresponding implementations of it are moving in this direction, not just for educational organizations but for any kind of organisation, and especially those that are multi national and globally distributed. It is now accepted that a well-established directory protocol offers a stable and tested environment for the purposes of storing and manipulating, among other things, student records. A scenario of implementation of the various learner specifications, which would not exclude the existing infrastructure of the learning institutions, could base the development of specifications on existing directories and corresponding protocols.

The most popular directory protocols, X.500 (OSI, 1993) and LDAP (Lightweight Directory Access Protocol) are able to support complex User Profile management scenarios. They predefine a set of classes, with a scope ranging from humans, through white pages, yellow pages to network devices (Wahl, 1996; Wahl et al., 1997). Taking into account that the most widely implemented directory services protocol is nowadays LDAP, the GESTALT project designed and developed its User Profile schema as part of an LDAP directory schema. LDAP is an extensible, vendor-independent, network protocol standard and supports hardware, software, and network heterogeneity. An LDAP-based directory supports any type of data and can be configured to play various roles while directly supporting various forms of strong security (authentication, privacy, and integrity) technologies. It is foreseen that general-purpose directory technology, such as LDAP, may be used to glue together disparate facets of cyberspace.

Recognizing the fact the metadata concepts and technology are going to play a leading role in building future directory applications, efforts now are directed on the interoperability of the two technologies, structured metadata and directory services (Tauber et al., 1999). The development of a standard and an underlying framework for the interoperability of the technologies will allow the semantically sound communication among these components (directories and metadata stores), resulting in sharing information held in what was previously considered to be different kinds of information stores.

CURRENT TRENDS IN METADATA AND PROFILES IN LEARNING TECHNOLOGY SYSTEMS

In the learning technology area, as already mentioned, the two main objects to be described are the learning materials and the learners. The learning resources are the digital informational content to be provided to the students. In this field, there are some very advanced efforts on the design of metadata. The Ariadne project (ARIADNE, 1999) pioneered the research in the field and later worked with the IEEE LTSC (IEEE LTSC, 2000:1) and IMS project (IMS, 1999:1; IMS, 1999:2). The GESTALT project has based its work on the LOM using it as a basic schema for reference while actively contributing to it. As previously noted LOM's design was based on the possible uses of the learning resources as part of learning systems. In this way noteworthy categories of the LOM schema are the Educational category that 'groups the educational and pedagogic features of the resource', and the Rights

category that "groups the features that deal with the conditions of use for the resource" (IEEE LTSC, 2000:1). On the other hand analogous efforts for the definition of the learner profiles in IEEE have been initiated and a preliminary work has produced the PAPI specification. The PAPI work divides learner information into four types: *personal information*, such as names, phone numbers, addresses, etc.; *preference information*, such as whether the learner prefers video over just plain text; *performance information*, this refers to the learners achievements in the context of learning environments (digitised or not) and finally, *portfolio information*, referring to the record of student achievements in terms of skills acquired, and abilities.

If the characteristics of the two specifications were compared, certain similarities can be seen. However, apart from the obvious analogy between what in LOM is called 'technical category' of the metadata structure and what in PAPI is called 'preferences information', which although not deriving from the same design rationale, are both 'translated' into primarily technical delivery options, the rest of the information is described in different terms. This is logical because of the different nature of the objects. LOM describes digitised information and PAPI a human educational profile. In spite of this, the fact remains, that a learning technology system needs to match the information held for these objects to construct efficient learning programs for the learners.

This aspect is further clarified when the IEEE Learning Technology Systems Architecture (LTSA) is examined. These objects are functionally positioned in a distance-learning environment and the relation between them is displayed. LTSA (Farance, 1999) defines a component architecture comprised by four processes: the learner entity, the evaluation, the coach and the delivery; a set of flows: learning preferences, behaviour, assessment information, performance information, query, catalog info, locator, learning content, multimedia, interaction context; and two stores: learner records and learning resources. The LTSA system components are depicted in Figure 1. The components that are of special interest are the learner records and learning resources stores and the coach process. The coach is acting as a mediator between the two stores (and the equivalent actors, like the teachers and learners) and one of its basic tasks, among others, is to match the information deriving from the user behaviour in the system, ranging from history to preferences, with the information describing the learning resources. This functionality of the Coach has been undertaken by the RDS component in the GESTALT project and is described in the next section.
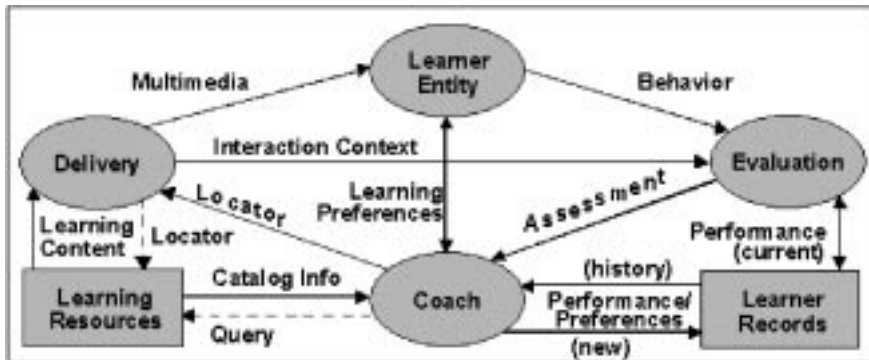
Fig. 1.   The LTSA system components.

As the education process in LTSA is defined in terms of a typical Information System, it follows that the processes may be enhanced, or made more intelligent, by quantitatively or qualitatively changing the flows. This could be done by choosing more efficient structures in the stores or if needed by introducing new stores.

As shown in Figure 1 the stores are accessed by more than one process. What is not so evident, but can be understood, is that the communication between the processes and the stores is very frequent. Further to this, it can be assumed that in the future (a) there may be more processes accessing the same components, and (b) there will be a need for interoperability between the different learning systems for exchanging data of both their stores (learning resources and learner records).

While common technologies for the manipulation of the learning resources and learner records stores — most usually 'home-grown' relational databases — can be a way to develop the depicted system, future needs will most certainly be hampered by scalability, transferability and interoperability issues. With regard to the stores, the following questions will need to be answered.

(a) How can the replication of the learning resources across different locations be tackled?
(b) How can the problems of transferring, renaming or deleting learning resources from the equivalent store be overcome without seriously affecting the system?
(c) How can the learner be guaranteed global access?
(d) How is the information contained in different stores going to be correlated?

Currently, researchers from various disciplines working on deployment issues of distributed services are confronting these questions. While learning technology area researchers (Forte et al., 1999; Rehak, 1997) and commercial products (Smith, 2000; http://www.mindlever.com/, http://www.saba.com/, http://www.smartforce.com/, etc.) are actively engaged in finding solutions to these issues with the available technology, it is also likely, that some new approaches will become available with from other disciplines also concerned with scaling up of systems, such as those looking into issues for areas other than educational (Bakker et al., 1999; Hoschek et al., 1999; Aiken et al., 2000).

For the purposes of the discussion here, it is clear that whatever the underlying technology, the two stores need to have the following characteristics.

(a) Learning resources should be stored in a way that makes possible the easy construction of retrieval mechanisms.
(b) Learning resources should be transferable at low cost.
(c) Learner records should be accessible from a range of components ranging from an institutional administration system to mail client programs.
(d) Learners need to have a suitable identifier acting as a globally unique identity.
(e) The properties of the different objects should be described in a semantically common language, i.e. equivalent properties must have qualitatively equivalent characteristics in their definitions.

In the next section the work undertaken in the GESTALT project, which tackled parts of these problems with the use of metadata and directory services is described.

THE GESTALT CASE

The implementation of a RDS with retrieval mechanisms on metadata combined with portable student records using a Directory Services model was undertaken within the framework of the GESTALT project (Wade et al., 1999). In this section, the overall GESTALT architecture is first presented, in order to describe a fully functional educational system and position the RDS and directory services in such a context. Then, RDS functionality and the way that metadata and User Profiles were used to enhance the retrieval efficiency of learning resources are described.

General Architecture

The objective of GESTALT was to design and implement an open on-line learning environment by bringing together and enhancing the results of previous projects and creating an integrated trial system, which encompasses the whole of the delivery of on-line learning from searching for learning courses and resources, enrolling on and following programs of study delivered over networks, and providing mechanisms for passing information between various parts of the system in a secure and trusted manner. These projects, RENAISSANCE (Konstantopoulos et al., 1998), PROSPECT (Wagner et al., 1997), and GAIA (Hands et al., 1999), looked at computer-based vocational training; quality of service over networks; and generic brokerage architectures, respectively.

As shown in Figure 2 the overall functional system comprised the discrete components described below:

Web Client. It is assumed that all User Services (i.e. all services available from GESTALT to the end users, as opposed to management functions performed by internal staff) are delivered to the desktop, using thin client (i.e. World Wide Web) technology. As part of the selection action the user has access to the RDS component. As part of the Instruction action the user has access to the Learning Environment functional component.
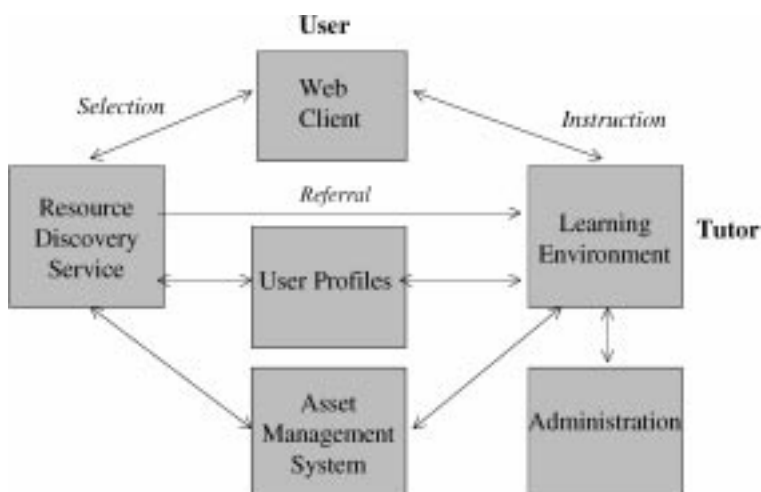

Fig. 2.   The overall GESTALT functional architecture.

Resource Discovery Services (RDS) allows users to explore what courses and modules are available from which institutions. The Resource Discovery Service functional component is based on the brokerage concepts and architecture developed within the GAIA project. As part of the Referral action the RDS makes available to the Learning environment (LE) details of the location of the selected educational resources.

Learning Environment (LE) utilizes an internal LE database to manage access and on-line study program progress. The goals of the LE are following.

- Provide holistic on-line support for the process of learning. This includes efficient and effective delivery to the desktop of on-line learning resources, over a variety of networks.
- Track individual student progress and achievements, enabling tutors to effectively support individual students, and achievement to be recorded.
- Support a flexible, modular curriculum.
- Manage on-line learners, in an organizational context.

User Profiles hold user preferences for LE interaction (e.g. preferred language), Resource discovery settings (e.g., frequently searched hosts/services), and other information such as personal portfolio and performance data. User profiles, when implemented with a directory service, act as a central point of data exchange between the different components. The User Profiles are used by the RDS and the LE. They can be maintained by the administration component or directly by the user (learner). Although, not shown in the figure, other components like the RDS, the LE and the administration component have or can have their own directories holding information specific for their functions but not of direct interest to the other components. For example, the LE can hold complete transaction records for each learner down to the grade achieved for each assessment whilst it publicizes to the other components only the exam results of the learner. Thus, the User Profiles component is not intended to substitute the other directories but to eliminate the need for developing separate systems interfaces for each data exchange between components.

The Asset Management System controls access to added value resources, for example, course/module data, which could be accessed at an extra cost program of study. This subsystem also publicizes learning objects (i.e. course module data) to the Resource Discovery Services.

The Administration Management Information System (MIS) supports back-of-house management of the training institution, in particular, relating

learning outcomes and student progress as assessed by the Learning Environment to learning objectives held in the MIS.

While the overall GESTALT architecture whose components are described above encompassed a wide view of an educational information system, the areas that are of concern to the work here are on the RDS functions and the use of User Profile for the retrieval of resources.

The RDS Architecture
The purpose of the RDS (sometimes referred to as the Broker) is to help learners locate Education Courses and Resources. The RDS is an integrated environment with an initial registration procedure of a learner (in this context client) and facilities for searching, previewing, and ordering Educational Courses and Resources. In order to do this more effectively the RDS, with the clients' approval, keeps track of a customers educational history and information about his/her computer and network connection.
Architecturally the RDS uses a three-tiered approach. The major components are the following.

- An Internet-based Client.
- An Education Search component, which keeps a cache of Course and Resource metadata and provides search facilities.
- A customer profiles component, which keep track of User Profiles by interacting with Trusted Education Systems and directly with users.

The RDS uses a CORBA-compliant Broker accessed via a web gateway. The RDS allows users to explore what courses and modules are available from which institutions. In addition a separate Functional Unit provides access to the GESTALT Learning Object Metadata. Protocol converters are provided so that the RDS can access metadata from services offering interfaces which are not CORBA based.

The resource discovery service functional component is based on the brokerage concepts and architecture developed within the GAIA project (Hands et al., 1999). The following Figure 3 summarizes the integration data model relationship between the RDS and the other system components.

As shown in Figure 3, the RDS exchanges metadata information with the Asset Management System using the GESTALT learning object metadata (LOM). User profile information is exchanged via the LDAP-based directory service using the PAPI data model. The RDS uses a number of leading edge technologies including XML, CORBA, the CORBA Trader, Java, X500 and
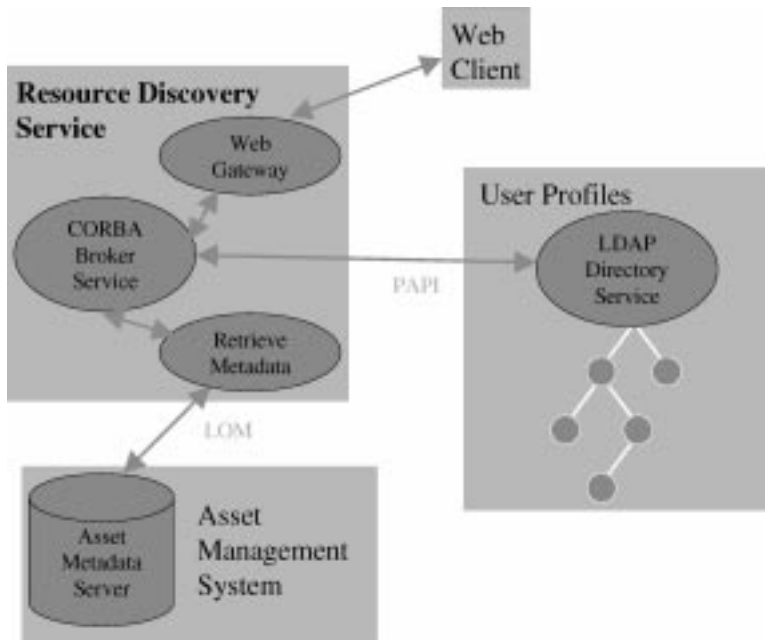
Fig. 3.   RDS functionality.

LDAP. Fundamental to the RDS design is XML-based educational metadata and LDAP user profiles.

The metadata used in the GESTALT demonstrator was based on the IEEE's Learning Technologies Standards Committee's (LTSC) Learning Object Metadata abstract data model (LOM) Version 2.5. From this specification an XML binding was developed to take into account the basic LOM data model as well as additional metadata attributes (extensions) identified during the early stages of courseware design. The main extensions included delivery format modifications, additions serving the Quality of Service concept support and a modification on the rights management support. This work is known as GESTALT extensions to metadata standards for education systems (GEM-STONES). The complete data model and bindings are described in detail in the public deliverable document D0401 "Courseware Metadata Design" (Foster et al., 1999).

The LDAP User Profiles were based on the PAPI specification, at least in regards to the information model. While the basic concepts of the profiles

derived from PAPI — the split of the profile in four distinct categories, personal information, preferences, performance information and portfolio with varying access rules- the implementation of PAPI in an LDAP server has rather radically changed the data model and has enhanced the functionality.

As can be seen in Figure 4, the profile has been split into objects that represent the basic entities, the security of which is determined by various access levels. Six classes have been constructed, which concentrate all the
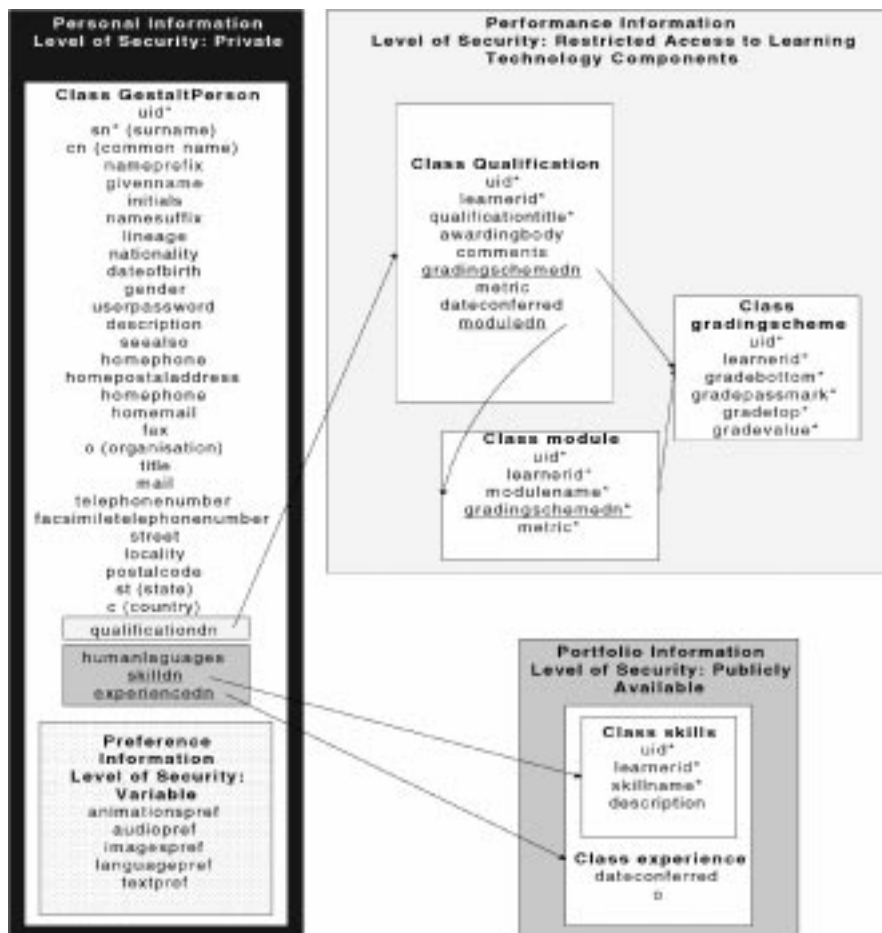

Fig. 4.   The GESTALT classes and their interconnections.

information specified by PAPI, as well as some extensions such as language preferences. These classes are: gestaltperson, qualification, module, gradingscheme, skill, experience. The classes are connected with distinguished names (dn) references. Each entry of the directory has a unique identity (uid) and a distinguished name (dn) that is globally unique conforming to LDAP architectural principles. The security requirements of PAPI are implemented through hierarchical access rules to the directory objects. The GESTALT demonstrator implemented a directory structure where the different objects of the directory were stored at different subtrees of the database. At each subtree an access rule was applied. In such a directory tree therefore, despite the fact that there are pointers connecting the objects, a logical and functional independence between the subtrees is preserved, permitting distribution of the profile parts amongst different operational departments of an educational organisation.

The Results of the GESTALT Demonstration

The GESTALT demonstration system provided the functionality of metadata-based educational resources retrieval. The GESTALT demonstrator permitted searches to be conducted on the learning material based on common metadata elements (e.g. title, author, etc.) or the full LOM structure.

The scenario around which the demonstration system has been built, is briefly described by the following three phases.

(a) A learner wishing to undertake some on-line training, contacts the RDS where he/she makes an initial registration through which a PAPI profile is collected.

(b) The learner can then search around a LOM described database of learning content either through a simple screen (which resembles to common library search utilities and where the LOM details are hidden) or through a thorough screen where he/she can browse the full LOM schema and actively choose the fields (tags) which will be part of his search query. This procedure can be executed on simple (where no profile information is included) or compound (the query is enhanced with elements from the PAPI profile) mode.

(c) When the suitable content has been located the learner may contact the learning service provider and follow its procedures (to access the content or follow a course).

Although, this procedure can be executed by a single person (usually the learner) the three phases can be completed partially or by different persons and roles. For example, an information scientist may only be interested in thoroughly searching and browsing content and service providers, or the initial registration may automatically be executed in the context of an agreement between the RDS and an organisational information system.

What was concluded from the GESTALT demonstrations was that while metadata could enhance the retrieval precision, the efficient exploitation of such a technology presupposed a user familiarity with the concept of metadata and the characteristics of the computational platform used. For example, a user should be familiar with the structure of the schema and the capabilities it provides in order to construct an efficient query. While this can be possible in common and well-known metadata schemes for most users (for example most bibliographic searchers can be considered familiar with the Dublin Core schema), when coming to complex, domain specific schemes (i.e. the full LOM structure), the familiarity cannot be presupposed.

The use of profiles was a determining factor in augmenting retrieval precision by matching metadata with the user performance, portfolio and preferred technology information. In contradiction with the 'simple search' where only content metadata information was used, the 'complex search' was 'silently' using the User Profile to enhance the performance of a query. The use of a profile for the selection of the right educational content for a learner was decisive in the way that it provided default (for the content) or preferable (for the learner) values in the content-learner matching, thus reducing informational retrieval noise. For example, if a learner (or the RDS acting for him) was searching by title educational materials, his profile character-istics, such as the technical capabilities of his machine, or the preferable reading language could be considered part of the query (corresponding of course to the equivalent parts of the metadata schema). This way even in the case that a user was not familiar with an educational metadata schema or was just not willing to add much information in a query, there was an option to enhance the query with the profile information.

Conceptually the above analysis is represented in Figure 5. At this figure the x-axis represents the metadata overhead on the information (i.e. how much descriptive — additional — information is provided for the data) and the y-axis represents the information retrieval precision as the proportion of retrieved material that is relevant to the user query. The main advantage of the complex search is near the beginning of the x-axis (that is where the
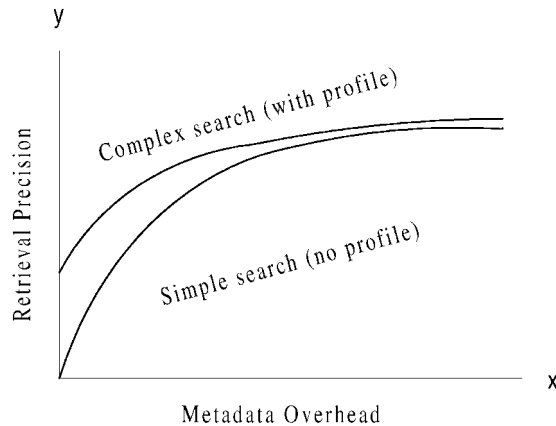
Fig. 5.   Complex and simple search.

metadata overhead is low). As metadata information is added in the search the
two curves converge. This is the result of the intentional (manual) use of the
profile information as part of the query by the user. On the other side when the
user does not insert much information in the query (which is the most usual
case) the complex search significantly improves functionality. It is also
apparent that the more accurate and up to date the information in the profile
is, the higher the starting point of the complex search curve on the y-axis.
Thus, the profile information in this case can be considered part of a user
query. This fact is clear when constructing an agent acting on behalf of the
learner. This agent is typically only based on profile information used to
search and retrieve information. If we expected an agent to produce rewarding
results searching a structured (described with metadata) information pool, it
would need to be 'fed' with suitable profile information.

Thus, what is generally expected in such cases is that the profile informa-
tion is comparable with the metadata information. But this does not always
happen. If for example, the technical requirements of a GEMSTONES
(extension of the LOM) educational content description is compared with
the LDAP description of the preferences (equivalent to technical requirements
in this case) part of a learner profile (implementing PAPI), it can be seen that
there are significant differences. In Table 1 the XML DTD description of this
part of GEMSTONES is compared with an excerpt of the definition of the
gestaltperson class in LDAP. Given that the RDS or any other component is
supposed to match content with learners it can only base its functionality on an

Table 1. Content Technical Requirements and User Technical Preferences.

GEMSTONES	LDAP-PAPI
<!ELEMENT Technical (Format,Size?,LocSpec+,Requirements+,InstallationRemarks*, OtherPlatformRequirements*,Duration?,OnLine?,Variant*)>	objectclass gestaltperson oid gestaltperson-oid superior top requires sn, uid
<!ELEMENT Requirements%RequirementsScheme;> <!ENTITY % RequirementsScheme "(Type,Name,MinimumVersion?,MaximumVersion?)">	allows . . . animationspref, audiopref,
<!ATTLIST Type type_requirement (OperatingSystem | Browser | WordProcessor | VideoPlayer | AudioPlayer | GraphicsPackage | InputDevice | OutputDevice | Processor | Protocol | Memory | StorageDevice | NetworkConnection) #REQUIRED>	imagespref, languagepref, textpref,

unambiguous association between the two descriptions. As can be seen, this is not the case. Apart from the unavoidable conversion from one way of description to another, which is already the subject of ongoing work (Tauber et al., 1999), the two descriptions remain qualitatively and quantitatively different. This is in spite of the fact that one would expect the two descriptions to have a minimum of differences at the technical level.

CONCLUSIONS AND FUTURE WORK

Metadata on the educational content can be a way to reduce the informational noise on the procedure of matching a learner's educational needs with this content. The learner profiles are another technology that, in addition to other functions, also helps to tackle the same issue. Although, it cannot be claimed that metadata and User Profiles can facilitate the way learning actually occurs, it can be claimed that under certain conditions these technologies can be used to add value to certain functions of the educational process. The task of accessing knowledge customized to the user is optimised by reducing the effort needed to locate the appropriate information, assuring that the information is in the suitable form and adapting to learner's knowledge evolution and learning focus.

The work described here shows that while the construction of a metadata infrastructure is essential for the future deployment of enhanced services in the educational sector, the User Profiles and the underlying directory services technology are the basis for the efficient exploitation of metadata. As one of the most critical functionalities of a learning technology system, that is, locating the suitable educational content for a learner, the User Profile values can act as supporting input to the queries performed by a broker or the learner himself. While this information is more than beneficial when a learner by himself is searching for content, it becomes critical when an agent acting on behalf of the user is executing the same task. Thus, the use of both the technologies can dramatically contribute to the precision of content delivery.

While the GESTALT project demonstrated the benefits from the implementation and use of the emerging standards in the field of educational metadata and User Profiles, it also displayed the inconsistencies of the proposed standards that have not been designed so that there is direct correspondence between their related elements. Thus, although there is no doubt that educational metadata and User Profiles do have some positive effects in the delivery of educational content, more work and research is required in the following aspects.

- Constructing finer metadata schemas and User Profiles specific for the educational area.
- Synchronizing the work and schemas in the critical fields of educational metadata and user profiles.

In addition, further work will be needed to accommodate issues of learning technology systems' adaptation to existing infrastructures and also to deal with globalisation issues on systems' accessibility and security.

In using the metadata and directories as the underlying technologies, new standards and tools will be required and perhaps even some rethinking of the context in which these are applicable. The IEEE's LOM is a specification that reflects the need of adding 'area of exploitation' specific information to the metadata. The PAPI specification is the equivalent effort — although still preliminary — on the User Profile field. It seems apparent that the added value will be based on products that will exploit these and other available specifications and technologies to provide to the learners with new products.

In conclusion, the work described here and future directions could be summarized as pertaining to the development of educational portals. These portals will try to exploit material that is described by metadata and will

enhance user capabilities for retrieving resources by permitting structured searches while benefiting from user profiling. More advanced services offered will include profile based agents executing search and notification tasks while in parallel a learner profile update mechanism will ensure that such a service will guarantee enduring functionality. Moreover the transformation of the learning procedure and the notion of learning as distinguished from education, provides the space to these portals to feature more generic information brokerage services. The clientele thus of such information may vary from final users, resellers and service providers. The GESTALT work is being carried on in this vein in the EASEL project, where the intended users of the system are educators seeking to reuse educational materials available on line and using the RDS to do so (EC, 1999). Another continuation of the GESTALT is in the GUARDIANS project (Guardians, 2000), where the idea is to mix both different devices (web TV, pcs) and networks, and yet use a brokerage structure to manage access to the information.

Finally, it is certain that learners will avail themselves of various resources located on different learning technology systems. It is also evident that in such a scenario a learner's learning achievements should be somehow captured and translated into profile information. This procedure should be independent from the learning technology system and the User Profile locations. These are then future services which will benefit from the developments of directory services and metadata infrastructures and most of all by their correlation.

ACKNOWLEDGEMENTS

REFERENCES

Aiken, B., Strassner, J., Carpenter, B., Foster, I., Lynch, C., Mambretti, J., Moore, R., & Teitelbaum, B. (2000). *Network policy and services: A Report of a workshop on Middleware*. Request for Comments 2768. February 2000.
ARIADNE. (1999). *Educational Metadata Recommendation, Version 3.0.* © *EPFL (Lausanne, Switzerland), KUL (Leuven, Belgium) and the ARIADNE Project*. (http://ariadne.unil.ch/Metadata/ariadne_metadata_v3final1.htm)

Bakker, A., van Steen, M., & Tanenbaum, A. S. (1999). From remote object to physically distributed objects. In *Proceedings of the 7th IEEE Workshop on Future Trends of Distributed Computing Systems*. South Africa: Cape Town.

BED Web Team. (1999). *Active Directory Architecture. White Paper.* (http://www.microsoft.com/WINDOWS2000/library/howitworks/activedirectory/adarch.asp)

Berners Lee, T. (1997). *Metadata Architecture, Axioms of Web Architecture*. (http://www.w3.org/DesignIssues/Metadata.html)

Buneman, P. (1997). Semistructured data. In *Tutorial in Proceedings of the 16th ACM Symposium on Principles of database systems*.

CDLP, (1999). *The California Distance Learning Project. What is distance education*? (http://www.otan.dni.us/cdlp/distance/whatis.html)

Darzentas, J.S. (1999). Sharing metadata: enabling on-line information provision. *OCLC systems and services*, *15(4)*, 172–178.

Dempsey, L., & Heery, R. (1997). *Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats*. Desire — development of a European Service for Information on Research and Education, Deliverable 3.2

Farance, F., & Schoening, J. (1998). *Public and Private Information (PAPI) Specification*. (http://edutool.com/papi)

European Commision. (1999). EASEL IST-1999-10051, *Educator access to services in the electronic landscape.* (http://www.cordis.lu/ist/projects/99-10051.htm)

Farance, F. (1999). *Learning Technology Systems Architecture (LTSA) — Specification Draft*. (http://edutool.com/ltsa/)

Forte, E., Haenni, F., Warkentyne, K., Duval, E., Cardinaels, K., Vervaet, E., Hendrikx, K., Wentland-Forte, M., & Simillion, F. (1999). Semantic and pedagogic interoperability mechanisms in the ariadne educational repository. *ACM. SIGMOD Record*, *28*(1), 20–25.

Foster, P., Kraner, M., & Graziano, A. (1999). D0401 *Courseware Metadata Design*. (http://www.fdgroup.co.uk/gestalt/D0401.pdf)

Grotschel, M., & Lugger, J. (1998). *Scientific information systems and metadata*. Sc October Takustraíe 7, D-14195 Berlin-Dahlem Germany Konrad-Zuse-Zentrum fur Information-stechnik Berlin. (Preprint)

Guardians, (2000). *Guardians Project Summary.* (http://www.fdgroup.com/guardians/)

Hands, J., Darzentas, J., Darzentas, J.S., Spyrou, T., Koutsabasis, P., & Smith, R. (1999). GAIA: a generic architecture and applicable platform for information brokerage. *Computer Communications Journal*. Elsevier, in press.

Hoschek, W., Jaen-Martinez, J., Samar, A., Stockinger, H., & Stockinger, K. (2000). Data Management in an international data grid project, In *Proceedings of the first IEEE/ACM International Workshop on Grid Computing*. Germany: Springer.

IEEE LTSC (2000:1). *Draft Standard for Learning Object Metadata. P1484.12/D4.0.* (http://ltsc.ieee.org/doc/wg12/LOM_WD4.htm)

IEEE LTSC (2000:2) P1484.13 *IEEE LTSC Student Identifier Working Group. Scope and Purpose.* (http://ltsc.ieee.org/wg13/index.html#S&P)

IMS (1999:1). *Learning Resource Meta-data Information Model, Version 1.0*. (http://www.imsproject.org/metadata/mdinfo01.html)

IMS (1999:2). *Meta-data best practice and implementation guide, Version 1*. Educause.

Koch, T. (1997). *The role of classification schemes in Internet resource description and discovery*. DESIRE — development of a European Service for Information on Research and Education. Deliverable 3.2.3.

Konstantopoulos, M., Spyrou, T., & Darzentas, J. (1998). A virtual vocational training environment: the Renaissance experience at the University of the Aegean. In *Proceedings of the 12th National Hellenic Operational Research Society Conference.* 3–6 September 1998, Pythagorion, Samos, Greece, in press.

Konstantopoulos, M., Spyrou, T., Koutsabasis, P., Darzentas, J.S., Lambrinoudakis, C., & Darzentas, J. (2000). The role of directory services in on-line education delivery: the GESTALT Example. *International Journal of Continuing Engineering Education and Life-Long Learning*.

Lawrence, S., Bollacker, K., & Lee Giles, C. (1999). Indexing and retrieval of scientific literature. In *Proceedings of the Eighth International Conference on Information and Knowledge Management, CIKM 99*, Kansas City, MO, 2–6 November, 139–146.

Microsoft. (1999). *Microsoft Windows Active Directory: An Introduction to the Next Generation Directory Services*. White Paper.

Microsoft. (2000). *Microsoft Metadirectory Services. Concepts and Architecture. Version 1.0. White Paper*. (http://www.microsoft.com/windows2000/library/howitworks/activedirectory/MMSintro.asp)

OSI, ISO/IEC 9594-1, (1993). *The Directory: Overview of concepts, models and services*. Recommendation X.500.

Rehak, D. (1997). *Carnegie Mellon Online: System Architecture*. (http://online.web.cmu.edu/public/information/publications/architecture/index.html)

Ruggles, R. (1997). *Knowledge tools: Using technology to manage knowledge better*. (http://businessinnovation.ey.com/mko/html/toolsrr.html)

Smith, B. (2000). *Higher Education & Enterprise Learning Management Systems*. (http://www.convergemag.com/Publications/CNVGNov98/highered/highered.shtm)

Steinberg, S.G. (1996). *Seek and ye shall find (maybe)*. URL: (http://www.hotwired.com/wired/4.05/features/indexweb.html)

SUN^TM (2000). *The Solaris^TM Operating Environment. Data Sheet*. (http://www.sun.com/software/solaris/ds/ds-sol8oe/)

Tauber, J., Hay, T., Beauvais, T., Burati, M., & Roberts, A. (1999). *Directory Services Markup Language (DSML)*. (http://www.dsml.org)

Wade, V., Riley, K., Banks, B., Foster, P., Evans-Mudie, N., Nicol, Y., & Doherty, P. (1999). *Object (Interface) Specification*. Work Package 5, The GESTALT project, final draft.

Wagner, M., Bøegh, J., & Krebs, A.M., (November 1997). *Web-based Virtual Learning Environments: Experiences and Futures*, WebNet, Toronto, Canada.

W3C (2000). *World Wide Web Consortium Issues XHTML 1.0 as a Recommendation*. (http://www.w3.org/2000/01/xhtml-pressrelease.html.en)