Sharing and Enriching Metadata in Architectural Repositories

Ann Heylighen[1], Herman Neuckermans[2], Martin Wolpers[3], Mathias Casaer[4], Erik Duval[5]
[1,2,4]K.U.Leuven, Dept. of Architecture, Urbanism & Planning, Belgium [3,5]K.U.Leuven, Dept. of Computer Science, Belgium
http://www2.asro.kuleuven.be/home/heylighen
ann.heylighen@asro.kuleuven.be

All over the world, students and teachers in architecture have been developing learning materials in various digital formats. Unfortunately, the material is not shared across school boundaries, and thus not exploited to its full extent. Mostly, technical and organisational limitations hamper the sharing and exchange of learning material, although this would benefit the global community of students and teachers in architecture. This paper presents a recently launched EU-initiative called MACE—Metadata for Architectural Contents in Europe—which aims at creating a European-wide space for the electronic descriptions of architectural information to be used in architectural education. The idea is to exchange and enhance the metadata of as many as possible digital repositories in order to allow searches by distant partners. Real access conditions to the data still remain those specific for each repository. By describing and discussing this initiative in its early stage, the paper aims to benefit from the exchange of ideas and experiences with similar initiatives, and to trigger the interest of new repository owners to join MACE.

***Keywords**: Digital repositories; metadata; architecture.*

1. Problem statement

Across the world, architecture schools have moved into the digital era. Students produce digital models and analyses of exemplary buildings; teachers develop digital learning materials, tutorials, PowerPoint presentations, animations etc. As a result, many schools have a large number of digital learning contents related to architecture. Unfortunately, the material is not shared across school boundaries, and thus not exploited to its full extent. Mostly, technical, organizational and perhaps even psychological limitations hamper the exchange of learning material, although this would benefit the global community of students and teachers in architecture.

The paper presents a European initiative that aims to simplify the sharing of digital learning materials across architecture schools in Europe. By enabling central access to large collections of architectural learning material, potential providers are encouraged to create and make available new learning resources, thereby gaining significant visibility and international recognition.

Basically the idea is to network already existing repositories and to allow distant searches in these learning materials by enriching their metadata. Every content/repository owner stays in control of his/her own learning material, but allows that the respective metadata are accessed and enhanced using advanced networking technologies.

After describing the general set-up and approach of the initiative (Section 2) and introducing, by way of example, one repository involved (Section 3), we will zoom in on the formalized description of learning objects through metadata (Section 4) and the infrastructure needed to search these metadata in a federated way (Section 5).

2. MACE

Aim and set-up

Metadata for Architectural Contents in Europe (MACE) is supported by the European eContentplus program (http://www.mace-project.eu). It aims at creating a European-wide space for the electronic descriptions of architectural information for use in higher education.

Throughout Europe and beyond, electronic resources, including learning materials for architectural education, are stored in many diverse and heterogeneous content repositories. The learning materials, in general called learning objects, are well used in the respective schools but most often focus on specific areas only, thus limiting themselves to information on specific knowledge domains. Contemporary architectural education requires a broad range of information, yet this information is usually stored in various repositories that cannot be accessed in a unified way. So far, no simple solution exists to find *all* suitable and overarching learning objects in architecture.

MACE addresses this gap by networking several already existing repositories and enriching the learning object descriptions to facilitate simple and efficient access to these materials. The MACE consortium involves ten partners from academia and industry. It builds on the experience gained in WINDS (Web based INtelligent Design tutoring System) (Kravcik and Specht, 2005), an e-learning platform containing 21 courses spread over Europe, and in the ARIADNE Foundation (Duval et al., 2001). Architectural content is provided by Iconda (Fraunhofer IRB), hosting 650.000 references and referencing 300 journals monthly, and DYNAMO (K.U.Leuven), addressed in Section 3, and complemented with 5000 learning objects from many universities worldwide through ARIADNE and the GLOBE network of learning object repositories.

Learning objects and metadata

Central to MACE is the use of metadata to provide machine readable descriptions of the learning resources to be made available and accessible. Learning resources are the content of the above mentioned repositories. Metadata can be defined as descriptive data, in this case about learning resources. In the context of MACE this means that each learning object (= data) is described through a set of properties and characteristics (= metadata). Each property is captured as a name/value pair in a metadata set describing the learning object. Name/value pair here refers to the name or identifier of the metadata field and its value. The value can be a simple type (string, character or number) or it can hold a reference to an ontology describing the property of the learning object in more detail.

In the MACE context, the set of available metadata fields describes the content, the domain, the necessary competencies for and the usage of each learning object. The content and domain metadata describe the nature of the learning object, its representation and/or manifestation in real world, its storage place and its location in some architectural classification scheme. Usage metadata describe how the learning object was used in which situations and contexts by whom.

The formalized description of learning objects through metadata serves a number of purposes, e.g. to facilitate more effective and efficient search and retrieval but also to enable more advanced user interfaces and learning strategies. This paper focuses on the usage of learning resource metadata for search and retrieval. In this context, search and retrieval refer to the retrieval of learning objects for specific applications, e.g. the targeted search for a certain learning object. These data provide highly valuable information on the usefulness of learning objects and thus can be used to improve the user-targeted retrieval of learning objects. Metadata describing the necessary competencies associated with each learning object enable further personalization services, e.g. by comparing the competencies required to those already acquired by a learner. The metadata on usage, competencies and context add additional descriptions about the learning resources in question, thus enriching the metadata provided by the content providers. The enriched metadata provide an improved information base on the available learning objects and therefore enable users to retrieve more precise and targeted learning objects for their context. A more detailed discussion of the metadata fields transcends the scope of this paper. The interested reader is referred to the related literature (Duval et al., 2002).

3. An example repository

Before going into the technical details of the MACE approach, this section introduces the DYNAMO repository, which will serve as illustrative example in the following sections. The Dynamic Architectural Memory Online (DYNAMO, http://dynamo.asro.kuleuven.be) is an interactive multimedia platform to share ideas, knowledge and insights in the form of concrete building projects among student and professional architects in different contexts and at different levels of expertise. It aims to incorporate quite literally the view of cognition underlying Case-Based Design and at the same time to extrapolate it beyond the individual (Heylighen and Neuckermans, 2000). From a technical point of view, DYNAMO can be thought of as a learning content management system (Heylighen et al., 2004). It is designed to support the creation, storage and (re)use of learning contents in the granular form of building projects. All learning contents are organized by a dynamic metadata classification system and stored in a data repository implemented in a relational database, subdivided into four sub-databases.

The *logon* database takes care of user privileges and administration, and distinguishes between three types of users. *Users* can consult the platform, but also feed it with new projects, project features and material. *Monitors* have extended privileges in that they can approve, alter or delete user added materials, and create new categories (see further). *Administrators* have access to all DYNAMO features, including user and monitor administration.

The *log* base logs all user interactions with the platform. For every interaction, it stores parameters such as the user identity (user name), the type of action performed (e.g. save file, view project, search etc.), its date and time, the location of the page registering the action, the query string used, and the client IP address. Data mining techniques process these logs to discover hidden facts about projects and their contents. Using a combination of statistical analysis and specific database queries, patterns and subtle relationships between the projects, metadata and other contents can be identified (Heylighen et al., 2006).

The *files* database contains all information about the material documenting the projects (file name, author, source, file type etc). Projects can be documented with a combination of various media: drawings (sketches, plans, sections, construction details), pictures, texts (in various languages), references (bibliographies, websites), digital 3D models, or analyses.

Last but not least, the *cases* database features a dynamic metadata structure. This structure is used to label each project by various features: project name, architect and location, but also aspects of form and space, function, construction and context. These metadata serve as filter criteria during retrieval and as links to related projects.

The metadata structure of DYNAMO serves as a template, enabling the hierarchical and flexible organization of project properties. For example, in order to specify the total estimated cost, a monitor merely has to add a 'cost'-metadata field through the web-interface to make this new category available for all projects.

DYNAMO's metadata are grouped into several 'windows', which enable users to approach and select projects from a specific point of view. The *ID* (identification) window comprises basic characteristics such as the architect's name, location and year of construction. The *design* window adopts a designer perspective and approaches projects through aspects of form and space, construction, function, context and concept, while the *theory* window enables theoreticians to select projects by tendency or movement. Recently a *construction* window was added, targeting users who learn or teach construction within the context of architectural education. Future addition of extra windows for other perspectives (e.g. history, conservation or reuse) is still possible. As we write, DYNAMO contains more than 600 architectural projects, documented by 9000+ files and serving about 1300 users from 12 different countries.

4. Mapping metadata

Having introduced an example repository, let us now discuss how MACE provides central access to learning objects in several architectural repositories across Europe.

The Learning Object Metadata standard (LOM) is used to formalize the description of the learning objects (Duval, 2005). LOM is an extensible IEEE standard which distinguishes between 9 metadata categories (general, lifecycle, meta-metadata, technical, educational, rights, relation, annotation, classification) and allows extension if deemed necessary. LOM incorporates and extends several fields of the Dublin Core metadata element set which is standardized as ISO Standard 15836-2003 (ISO DC). The 15 Dublin Core metadata elements are used to describe a resource in general, e.g. the title, creator, format etc. MACE starts from LOM using its XML binding to enable interoperability with repositories already using this widely recognized standard.

For DYNAMO, two types of learning objects can be distinguished. On the one hand, there are architectural projects as manifested in real world (whether or not effectively built), whose properties are described via domain metadata, such as the architect's name, location, building program etc. On the other hand, there are digital files documenting these projects (sketches, plans, texts, models etc.), for which technical properties like the file type, file name, author name, file size etc. are captured. Within MACE, the metadata of both learning object types are mapped onto the LOM standard based on the application profile, i.e. a specifically developed and defined set of metadata fields. Mapping is straightforward by including the content repositories directly or referring to suitable domain ontologies where possible. An excerpt of a LOM metadata set in XML notation illustrates how parts of a DYNAMO *project*—the Musée des Beaux Arts in Lille—are mapped onto LOM:

```
<lom xsi:schemaLocation=" http://
ltsc.ieee.org/xsd/lomv1.0/lom.xsd">
  <general>
    <identifier>
      <catalog>DYNAMO</catalog>
      <entry>1<//entry>
    </identifier>
    <title>
      <string language="fr">Musée des
Beaux Arts Lille</string>
    </title>
  </general>
  <lifeCycle>
    <contribute>
      <role>
        <source>LOMv1.0</source>
```

```
          <value>author</value>
        </role>
        <entity><![CDATA[BEGIN:VCARD
                   FN:Jean-Marc Ibos
                   N:Ibos;Jean-Marc
                   VERSION:3.0
                   END:VCARD
        ]]></entity>
      </contribute>
      <contribute>
        <date>
          <dateTime>1997-01-
01T22:00:00.00Z</dateTime>
        </date>
      </contribute>
    </lifeCycle>
    <rights>
      <cost>
        <source>LOMv1.0</source>
        <value>yes</value>
      </cost>
    </rights>
    <classification>
      <taxonPath>
        <source>
          <string
language="en">Climate</string>
        </source>
        <taxon>
          <entry>
            <string
language="en">temperate</string>
          </entry>
        </taxon>
      </taxonPath>
    </classification>
</lom>
```

The excerpt shows how DYNAMO properties can be mapped onto respective LOM fields. For example, the <title> field captures the title of the learning object as well as the fact that the title is given in French (see the field <string language="fr"> which specifies a property language with the value fr for French). Apart from general information like the title, information about the author, c.q. designer is captured in the <lifecycle> tag. Here, the excerpt illustrates how other standards are reused in LOM, e.g. the vCard (vcard, 1998) standard to capture the author information. The <rights> tag (including the <cost> tag) captures information on the access and usage rights of this resource, here simply specifying that there are some costs involved. The <classification> tag shows how LOM includes information about the learning resource that is specified elsewhere in ontologies. Here, it specifies the climate in which the content of the learning resource is situated by specifying the location of the definition of the value as well as the value "temperate" itself. The LOM metadata set describing an image that represents the Musée des Beaux Arts does not include the above metadata. Instead, it describes the image itself (e.g. size, format etc.) and further contains a reference to the LOM metadata set representing the Musée des Beaux Arts.

Searching for the learning object from the example means to search through the metadata descriptions of all learning objects. For example, searching for learning objects with the title "Musée des Beaux Arts Lille" simply means to look at the <title> field of all descriptions. Finding the project representing the Musée enables the user to find all learning objects referencing this project as well.

5. The MACE infrastructure

The MACE infrastructure is based on a hybrid combination of harvesting metadata from and federating searches to existing content repositories. Existing learning object metadata are enriched with information on the use, the context and the competences of the user. The approach aims to make the learning objects in all repositories findable for the regular user, through a mechanism that allows simultaneous searching over the content of all federated content repositories. Each repository makes its learning objects accessible through the aforementioned metadata descriptions, e.g. in the case of DYNAMO through project and file categorizations. A major challenge to be addressed by the MACE project is that the organization of these descriptions differs across repositories, which hampers "semantic interoperability". In order to facilitate searching, the learning object metadata of each repository are harvested into one central content metadata repository. Access to the actual learning object remains controlled and managed by the content provider as only the metadata descriptions of learning objects are shared in the central metadata repository.

Harvesting here means transferring metadata from the respective repository into the central content metadata repository on a regular basis. Only the metadata describing the learning objects are transferred. The learning objects themselves stay in the repository and thus in control of their owner without changing the respective IPR. The metadata are harvested through harvesting interfaces at each content repository that implement the Open Archive Initiative Protocol for Managing Harvesting OAI-PMH (OAI, 2002). The central content metadata repository also offers an OAI-PMH interface so that interested content metadata providers can retrieve eventually enriched metadata suitable for their learning objects. Each repository therefore provides its information described with the required OAI-PMH metadata fields location, identifier and the last modification date, which are all mapped onto LOM. Furthermore, the repositories' internal metadata structure is mapped onto the LOM standard at the repository.

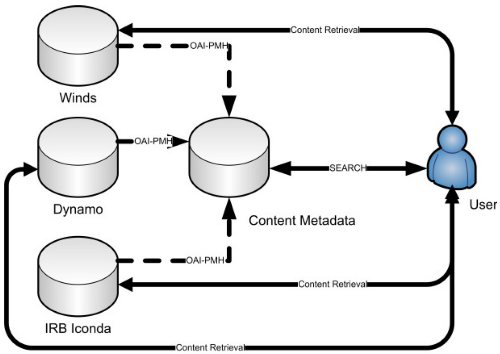

Figure 1
MACE harvesting infrastructure

Figure 1 shows the conceptual set-up of MACE. The metadata describing the learning resources provided by the architectural learning object repositories (DYNAMO, Iconda, and WINDS) are harvested through the OAI-PMH protocol into the MACE content metadata store. The metadata store supports the search facility through which the user finds references to available and suitable learning objects. In order to consult the learning object, the user accesses the learning resource directly at the provider.

DYNAMO is integrated into the MACE architecture by implementing the OAI-PMH interface. If users search for content related to a topic represented in DYNAMO, the metadata information in the central repository refers them to the DYNAMO repository, where repository-specific access control mechanisms manage access to the learning object.

MACE aims to provide personalized services, such as personalized search taking into account the search terms, user's context and usage behaviour. Such advanced services require vast amounts of information about the user, her context and the available learning resources. Therefore this information is captured in a number of federated stores: the metadata store (describing the learning objects), the contextual metadata store (describing the context in which learning objects were used) and the usage data store (which describes usage of learning objects).

A federated search application searches the metadata stored in the central metadata store to find suitable learning objects, eventually taking usage and contextual metadata into account. The federated search service is enabled through the Simple Query Interface (SQI, 2005). SQI allows for the federation of queries and the agglomeration of the query results. SQI can be combined with any query language. It is used, for example, in the GLOBE consortium to federate queries over the global network of learning repositories (Ternier et al., 2005). OAI-PMH is used to harvest the metadata and collect the usage data. The idea here is to harvest the log files from the various repositories, which enables the provider to control which information is exchanged with the usage data store.

In general, the MACE architecture enables the user to find learning resources in a large number of repositories while working with one search application. Furthermore, by using information about the context of users and the usage of learning objects, MACE enables a more personalized search experience with more suitable and better targeted results than offered when searching in the various repositories separately.

6. Summary and conclusion

This paper has presented the approach adopted in MACE, a European initiative to enable simple and personalized access to learning resources in architecture. MACE aims at sharing efforts, information and digital learning materials across different architecture schools. Basically, the idea is to network a number of already existing digital repositories in architecture, and to facilitate access to these learning materials by harvesting and enriching their metadata.

As we write, the MACE consortium is creating a first prototype, which will be revised and improved in subsequent iterations. At this point, it is obviously too early to assess the impact of a complex effort such as MACE, and to measure its added value compared to the services offered by individual repositories. By presenting the project in its early stage, however, we hope to benefit from the exchange of ideas and experiences with similar initiatives, and to trigger the interest of new content partners. MACE is continuously looking for new partners willing to open up their contents in return for an increased reach. Currently, more e-content owners willing to participate in the project are being scouted via the European Association of Architectural Education (EAAE, http://www.eaae.be), which is a major partner in MACE.

Acknowledgements

MACE is co-funded by the European eContentplus program, project nr. ECP-2005-EDU-038098.

References

Duval, E. (Ed.): 2002, IEEE Standard for Learning Object Metadata 1484.12.1-2002, IEEE 2002.

Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Van Durm, R., Hendrikx, K., Wentland-Forte, M., Ebel, N., Macowicz, M., Warkentyne, K., & Haenni, F.: 2001, The ARIADNE Knowledge Pool System, Communications of the ACM 44 (5), May 2001, pp. 73-78.

Duval, E., W. Hodgins, W., S. Sutton, S. & Weibel, S.: 2002, Metadata principles and practicalities, D-lib Magazine 8 (4), April 2002, pp. 1-16.

Heylighen, A. & Neuckermans, H.: 2000, DYNAMO: Dynamic Architectural Memory On-line, Educational Technology and Society, Vol.3, No.2, April 2000, pp. 86-95.

Heylighen, A., Neuckermans, H. & Casaer, M.: 2004, ICT revisited - from information & communication to integrating curricula?, ITcon Vol. 9, Special Issue Digital Media Libraries, 101-120 (http://www.itcon.org/2004/7).

Heylighen, A., Neuckermans, H. & Casaer, M.: 2006, UnaWare. Supporting tacit design knowledge exchange, International Journal of Web-Based Communities, 2(1), Jan 2006, pp.31-44.

ISO DC: ISO Standard 15836-2003 formalizing the Dublin Core Element Set.

Kravcik, M. and Specht, M.: 2004, Flexible navigation support in the WINDS learning environment for architecture and design, Proceedings of third international conference on adaptive hypermedia and adaptive web-based systems (AH 2004), Ed. Paul De Bra, Eindhoven, The Netherlands, August 23 - 26, 2004, pp. 156-165.

OAI: 2002, Open Archives Initiative Protocol for Metadata Harvesting, Protocol Version 2.0 of 2002-06-14.

SQI: 2005, European Committee for Standardization (CEN) Workshop Agreement on "A Simple Query

Interface Specification for Learning Repositories", November 2005.
Ternier, S., Olmedilla, D. and Duval, E.: 2005, Peer-to-Peer versus Federated Search: towards more Interoperable Learning Object Repositories, - ED-Media 2005, (Kommers, Piet and Richards, Griff, eds.), pp. 1421-1428.
Vcard: Proposed IETF standard RFC 2426, http://tools.ietf.org/html/rfc2426.