Developing a Network of Cultural Heritage Objects Repositories for Educational Purposes

Kerstin Schmidt, Hans-Christian Schmitz, Martin Wolpers

Fraunhofer Institute for Applied Information Technology FIT, Schloss Birlinghoven,
53754 Sankt Augustin, Germany
{kerstin.schmidt,hans-christian.schmitz,martin.wolpers
@fit.fraunhofer.de}

Abstract. Natural history museums collect and provide access to digital representations of artifacts from their vast collections. The representations are used, among others, to facilitate their use in educational settings. For example, school teachers use them to prepare class visits to the museums. In the context of the European project Natural Europe, these repositories are to be bridged in order to enhance the usage experience in learning scenarios. Based on a thorough requirements engineering process, respective architectures are defined and services provided. In this paper, we discuss the outcomes of the requirements engineering process. While teachers appreciate the provision of fairly complex learning paths through museums, they also need direct access to the individual learning resources aka artifact representations. Finally, the technical architecture of the Natural Europe system is presented as a result of the requirements engineering process.

Keywords: natural history, requirements engineering, learning repository, learning path, content harvesting, metadata

1 Introduction

Throughout the past decades, natural history museums (NHM) have undergone fundamental changes in means of storing information on the objects in their repositories, those are cultural heritage objects (CHO). With the development of information technology and data management systems, the museums applied new technologies for the creation and the maintenance of digital data collections. While digital storage facilitates the administration and accessibility of the data collections compared to the previous analog methods, they also imply great challenges to the architecture of the respective storage systems concerning their interoperability. Not only is there a need for those systems to cooperate NHM-internally, but also for the cooperation on higher, national and cross-national levels as demanded by modern formal and informal educational methods. The Natural Europe project aims to overcome these barriers by providing a coordinated solution on a European level,

which will connect the different NHMs' repositories and provide educational services on top of these federated repositories to enhance teaching and learning activities.
The following exemplary use case describes a scenario, where Natural Europe facilitates the utilization of cultural heritage contents for educational purposes: according to the curriculum of his school, a teacher wants to explain the concept of evolution to his pupils. He plans a visit to the natural history museum of Lisbon to vividly explain the idea of evolution and its implications. The teacher develops one or even several learning paths for the museum visit. For preparing the visit, he collects various materials on the topic, including information on the Archaeopteryx, a dinosaur found in Bavaria and considered to be the evolutionary link between dinosaurs and birds. The material on the Archaeopteryx is provided by the museum in Eichstätt and made accessible via Natural Europe. In a post-visit phase the pupils develop presentations reflecting the museum visit. To this end, they use the Natural Europe repositories as a source for their research.
To implement the Natural Europe objectives, several NHMs, pedagogues, educational technologists, metadata experts and user groups work on European-wide integration and standardization processes, bringing together high quality cultural heritage contents of European NHMs and from digital libraries like Europeana [1], in order to enhance natural history and environmental education. The technical realization of the project does not start from scratch, but rather uses valuable experiences from related European projects. An enhanced version of the tool for authoring metadata on multimedia files that has been developed in the DELOS project [2] is used in Natural Europe for authoring digital cultural heritage contents, and the tools for the design of learning activities – like the aforementioned pathways – are based on similar tools used in the LOGOS project [3] and the e.KnowNet Network of Knowledge [4]. For the search and navigation components the interface developed in the MACE project [5] for architectural contents is adapted to Natural Europe's cultural heritage and learning contents.

After a description of the project's context and its main objectives in section 2, section 3 focuses on the results of the technical and educational requirements elicitation processes, shaping the overall Natural Europe system architecture, which is be presented in section 4. Section 5 concludes the paper and addresses the next steps.

2 The Natural Europe Project

The Natural Europe project, fully titled "Natural History & Environmental Cultural Heritage in European Digital Libraries for Education" [6], is supported by the European Commission under the Information Communication Technologies Policy Support Programme[1]. It focuses on two main topics, namely the federation of the participating natural history museums' digital data collections on the repository level, and the design and implementation of educational services and tools consuming the contents of these repositories.

[1] http://ec.europa.eu/cip/ict-psp/index_en.htm

2.1 Federated Cultural Heritage Repositories

There are several difficulties to overcome in order to build up a repository available for higher level, Europe-wide services. These technological aspects result from the way the respective NHMs' data collections have been built up and have evolved over time. The process of creating the data collections – by digitization of analog contents or by direct creation of digital contents – has been implemented by each institution autonomously and according to its individual requirements, resulting in a large variety of different data collections of a different technological nature. This hinders the integration of digital data collections into commonly accessible repositories and thus vast amounts of contents remain unexploited. To overcome these restrictions, Natural Europe aims at federating the diverse data collections into one comprehensive repository, providing unified access to the contents originating from various sources. This enables the interaction with other CHO federations, like the Europeana digital library. Europeana collects digital contents originating from museums, libraries, archives and other audio-visual collections from all over Europe. The Natural Europe federation will serve as an aggregator for Europeana and vice versa, since both federations will consume contents of and upload contents to each other. Currently, six natural history museums are members of the Natural Europe consortium: the Arctic Centre in Rovaniemi, Finland [7], the Hungarian National History Museum in Budapest, Hungary [8], the Jura Museum in Eichstätt, Germany [9], the National Museum of Natural History in Lisbon, Portugal [10], the Natural History Museum of Crete, Greece [11] and the Estonian Museum of Natural History in Tallinn, Estonia[12].

2.2 Educational Services

Apart from preserving cultural heritage, one of the main interests of the NHMs is education. Analogous to the technical difficulties described before, learners and educators are facing similar problems. Usually the NHMs are connected to educational institutions and offer different educational services on a local level. Due to the lack of interoperable systems, they do not exploit the huge amount of information residing in other institutions.

Based on the federated repository developed in the context of Natural Europe, the project will develop educational tools, allowing the search, retrieval and creation of educational objects related with the contents of the cultural heritage repositories. The major objective from this educational point of view is the concept of learning paths, also called educational pathways. In this context, the term learning path refers to a sequence of learning activities. These are combined by educators in a certain order, ensuring that the learner increases his knowledge progressively by following the path. The educator is free to choose any structure of the learning path and its contents. While the term cultural heritage object refers to the digital objects in the NHM collections, the learning modules contained in a path as well as the path itself are called learning object (LO). These learning paths reflect the various usage scenarios of the cultural heritage contents to be implemented:

1. Educators want to enrich their curricula with structured or unstructured learning activities containing digital scientific material from the NHMs. Natural Europe will facilitate the access to LO repositories and also provide the tools for creating such learning paths.
2. Learners require access to rich scientific material to support their learning and research activities, and they need access to LO repositories (or parts of it) to follow the paths implemented by their educators.
3. Visitors want to enhance their visit by following structured activities through the NHM, enhance their experiences through pre- and post-visit activities.

In the following section we will describe the requirements for the project, reflecting the educators' concrete needs and serving as input for the design process of the system architecture.

3 Requirements

The technical and educational constraints have been set by classical methods of requirements engineering, such as workshops, interviews and questionnaires. After basic steps like setting the scope of the project and the identification of stakeholders, detailed analyses of the present situation in the NHMs nowadays have been made. Building on these, the explicit requirements for the features of future NHM portals have been elicited. In these processes representatives from the NHMs were involved, i.e. museum curators and collection experts from the particular departments in the NHMs for the technical part, as well as NHM-internal and external pedagogues of various educational levels for the educational part.

3.1 Content Requirements

The technical requirements mainly deal with the identification of the nature of existing collections in the NHMs participating in the project and the constraints these impose on the layout and implementation of the Natural Europe system.

It turned out that only few similarities could be found in the characteristics of the NHM collections, meaning that Natural Europe has to face the task of finding means of unifying the contents of the different collection for the use in a common repository.

Starting with data storage, many distinct methods have to be considered. In the relevant collections file-based storage systems can be found, like Microsoft Excel or plaintext XML files, as well as database management systems, like Microsoft Access, MySQL, Filemaker, and other third party content management systems. Apart from that, the data structures are also of different nature – while for many collections self-defined metadata schemas not conforming to any standards are used, others are compliant to common standards, e.g. Darwin Core[2], Dublin Core[3] or ABCD[4], and

[2] http://www.tdwg.org/activities/darwincore/
[3] http://dublincore.org/
[4] http://www.tdwg.org/activities/abcd/

some do not even have the option of providing collection metadata in XML format. The accessibility aspect also implies great difficulties on the Natural Europe project since the collections are far from being completely web accessible. Some of them are, but most of them hosted on NHM internal servers, reside on personal computers or are stored on DVDs.

3.2 Technical Requirements

Looking at the curators' requirements for the Natural Europe repositories, it can be said that they need means for uploading content to the federated repository at least in a semi-automatic way: due to the amounts of CHOs it would be impossible to enter the data manually into the repository and due to the different metadata formats on the NHM side the data transfer cannot be done automatically, since the data have to be transformed into a standardized Natural Europe format first. This standard must have general metadata fields for capturing information on the CHOs as required by the interoperability with Europeana, and possibly more fields to capture additional information coming from the nature of the CHO's collection. For the maintenance of the data, including the creation of digital CHOs in the federated repository as well as the modification or extensions at a later time, the curators require an authentication and authorization system to protect their data from unauthorized access, and also a rights management system for the CHOs. Generally, the Natural Europe Intellectual Property Rights (IPR) schema, which is to be developed, should be applied to the contents, but there are other cases in which the access could be restricted to certain user groups. Lastly, they asked for Web2.0 features for the discussion on the CHOs, e.g. commenting on the content, notifying the author of errors or changes, Wikipedia discussions and Amazon reviews have been mentioned as examples for that.

3.3 Pedagogical Requirements

Previously the targeted outcomes of Natural Europe have been summarized as educational tools allowing search, retrieval and creation of learning objects. The educational requirements are far more complex than that and will be described in the following paragraphs.

Educational Metadata. Firstly, the learning objects require general metadata, just as the CHOs described in the previous subsection do. Second, regarding the requirements for the actual LOs, one of the most required features is the adaptability to the local curricula. Since the latter vary not only between the types of educational institutions but also in the different European countries, classifications are needed, clearly stating for which purpose and target audience they are suited. These classifications include educational characteristics like the type of the educational activity, target audience, educational level, school type, topics/context, duration of the activity and difficulty.

These classifications are necessary for the usability of the LOs because only by having these educational metadata the user perform reasonable actions on the datasets, like browsing for a specific pathway e.g. an elementary school course in biology on vertebrates of the Stone Age, or a university lecture on geology etc.

Learning Objects. Through the various procedures of the requirements engineering it turned out that there is a general need for learning paths of different educational levels and for different target groups. Having a repository containing learning paths (LPs) enriched with content from NHMs all over Europe will facilitate the educators' tasks of creating attractive structured learning activities including pre- and post-visit activities. This will help the educators to raise the learners' interest in the respective topic, it will increase the reusability and the dissemination of the educational pathways, which would otherwise might only exist on paper in somebody's desk, and additionally, it will help educators who are not that familiar with the adaption of certain pedagogical models to apply the pathways in their teaching activities. Private visitors of the NHMs that are interested in the pathways may also use them to plan their visits of a certain museum.

Apart from the learning paths, there is also a need for less structured educational activities. In this case, 'less structured' means that educators require a repository of educational contents to not only provide pre-structured activities but also smaller learning modules, which are not bound to an explicit educational model. This allows them to implement complex learning paths as well as smaller educational activities. Finally, the need for unstructured learning activities may even be drilled down to the usage of the CHOs in learning activities because in many scenarios educators might only want to use information on just one single object to their own learning activities.

Platform Features. The requirements elicited so far can be summed up by the following basic features the Natural Europe platform should offer. Further requirements resulting from the discussions on these features during the requirements elicitation process will be listed as well.

- The creation of new learning paths: this rather general task requires the platform to offer features like a user management system for authentication and authorization issues, options to choose between various learning path templates to be adapted by the learning paths, a repository for the storage and access of learning paths and most importantly the connection to CHO portals for objects to be integrated in the pathways. Apart from the template selection and adaption tasks, these features also apply for the revising and editing of existing learning paths.
- The usage of existing learning paths: to enhance the usability of the repository, the Natural Europe platform requires features like searching for learning objects, the navigation through the federated repositories, access and digital rights management. Similar to the technical requirements for the CHO repository, educators also would like to have the features of commenting and contacting the author to discuss and enhance their respective contents.

After one year of the project, the system architecture, which is described in the next section, has been designed according to the given requirements. While for editorial work on CHOs and LOs there are already prototype tools being brought to use, the specification of external content integration is currently being finalized.

4 The Natural Europe Architecture

From a technical point of view, the requirements for the Natural Europe architecture, whose main components are shown in figure 3, can be grouped into the following three categories, representing the main target outcomes of the project:

- A cultural heritage object repository: this repository provides means for the integration and the maintenance of the digital data collections coming from various NHMs into one standardized Natural Europe repository. Also, it ensures the interoperability of the project repository with the Europeana digital library.
- An educational object repository: this repository provides means for hosting and maintaining the learning paths as well as the learning path templates, coming from the NHMs and the users of the Natural Europe portal. Beyond that it provides interfaces for exposing the educational contents to other learning object repositories.
- Graphical user interfaces (software and hardware) providing access to the platform services and allowing the application of the learning paths virtually or inside the NHMs.

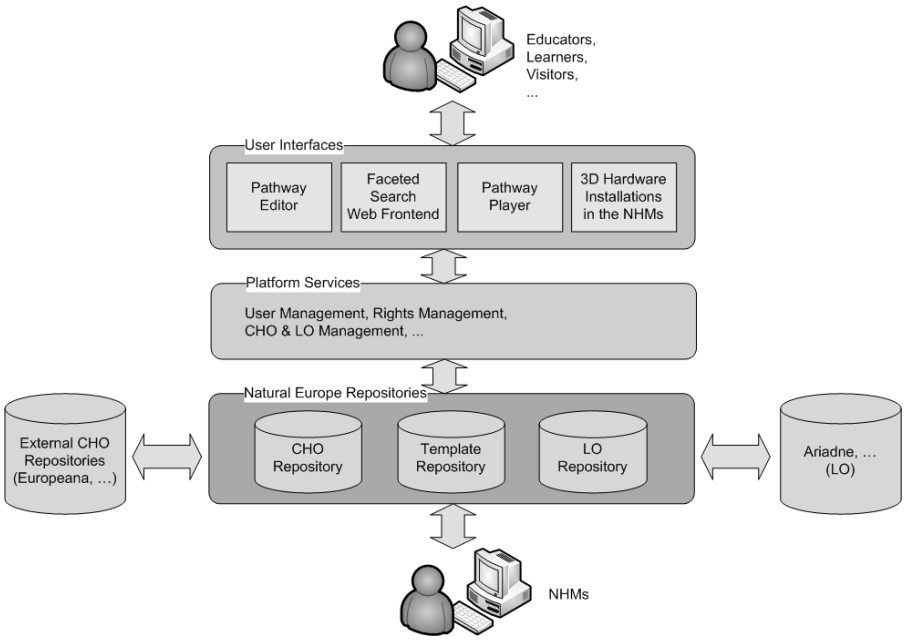

Fig. 1. Overview of the Natural Europe Architecture

4.1 Cultural Heritage Portal

The Natural Europe cultural heritage portal serves the purposes of a digital collection repository. While the NHMs are the main content producers for the repository, the project's learning objects portal, the Europeana digital library and other LO portals are the main content consumers of this portal.

Metadata Standards. To ensure the interoperability of the necessary software tools and components, common standards for the data formats and for the interfaces connecting the specific components have to be agreed upon.

After the evaluation of eligible standards, it was decided to use the Europeana metadata standard as the storage format cultural heritage object metadata in Natural Europe, since the exchange of data to and from Europeana, like harvesting, searching and retrieving, is a crucial requirement of the project. Thus the project supports the current Europeana metadata standard Europeana Semantic Elements (ESE) [13]. The ESE standard is a metadata standard based on Dublin Core [14], containing generic data fields for information on library, culture, museum and other archive objects.

The Europeana Data Model (EDM) [15] is a new approach of the Europeana federation which will support the full richness of the providers' metadata. It does not map the various standards to one central metadata schema, but it follows an open semantic web approach, by which the various metadata standards can be accommodated in an interoperable way. Natural Europe currently supports the ESE standard, but it will adapt EDM as soon as it is released by Europeana.

Portal Connectivity. To support the standardized metadata exchange between the Natural Europe repository and not only Europeana but also other CHO repositories, the project will make use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The OAI-PMH protocol [16] is used to exchange Dublin Core-based metadata of records stored in a digital archive between other archives.

Portal Services. The services of the CHO portal can be grouped together into repository, harvesting and administration services. The repository services provides basic features like storage and maintenance of the CHOs and indexing the stored data for search services. Furthermore it controls the access to the repository contents which will be protected depending on the users' rights and on the digital rights for the CHOs. It is also responsible for the necessary functions to produce and consume CHO information transferred to and from other repositories or services. The harvesting service manages and performs the actual harvesting processes, i.e. is the data import from participating institutions like the NHMs in the Natural Europe project. Finally the administration service implements features like the management of users and access rights and the administrative control of the other services.

Data Flow. The lifecycle of a cultural heritage object in the CHO portal of a NHM can be illustrated by the following steps of the data flow from the participating NHMs

to the common portal. First the NHMs will have to prepare the contents of their collections, which includes a) the creation of digital representations of the objects in case they are only present in an analog shape and b) the publishing of the contents, meaning that the objects have to be web-accessible for consumers like the users of the Natural Europe portal. Each participating NHM selects contents from its various collections and contributes the chosen CHOs to the project. For the use in the Natural Europe portal and others like Europeana the CHOs have to be supplied with persistent unique identifiers, ensuring that the original digital CHO can always be accessed via a URL that is not changing. Before the CHOs can be uploaded into the NHM's Natural Europe repository, their metadata has to be transformed into the common metadata schema which was determined to be the ESE/EDM schema. To do so, for each metadata schema which is used by the NHM a mapping has to be defined which is then used during the harvesting process to transform the original metadata into the ESE format – the EDM format will be used as soon as it's released by Europeana. Once the CHOs are imported to the repository, the NHM curators can administrate them, including modification or deletion as well as the enrichment with additional metadata. The step of mapping the metadata, importing the CHOs to the local NHM repository and the further administration and enrichment of the CHOs will be performed through a web-based authoring tool, which is already done by the museums. Finally the collected metadata in the NHM repositories are harvested into the common Natural Europe CHO repository which will provide the connection to and exploitation by other services and portals.

An overview of the components involved in the CHO portal, including the CHO data sources and the interaction with external portals, is shown in figure 2.

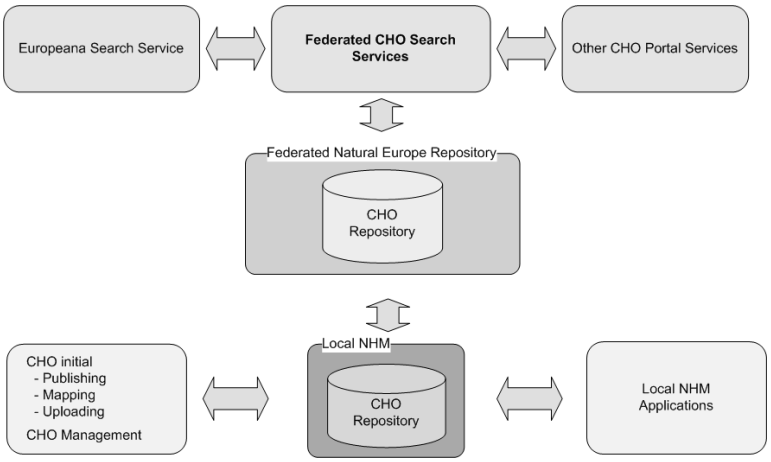

Fig. 2. The Natural Europe CHO Repository

4.2 Learning Object Portal

The second main target of the Natural Europe project is the creation of a learning object repository, hosting diverse learning activities which are enriched with contents from the internal cultural heritage object repository as well as external sources.

In contrast to the cultural heritage object repository, where the NHMs serve as main content producers, the learning object repository content will be produced by NHM educators, external educators (schools, universities, lifelong learning institutions etc), and other interested individuals. The content consumers on the other hand will be learners (pupils, students etc.) and also the educators (reusing existing contents) and other learning object federations.

Metadata Standards. In the context of the Natural Europe learning objects repository the Learning Objects Metadata (LOM) standard [17] will be used to describe the learning paths as well as the learning path templates. LOM is an internationally accepted standard published by IEEE for the description of learning objects, which can be any entity, digital or non-digital, that is used, re-used or referenced during technology-supported learning. This standard specifies a minimal set of fields necessary for the management of those learning objects, and at the same time provides a wide set of extensible data fields to adapt the standard to the needs of the usage scenario. The following categories of data elements are described by the LOM standard:

- General – general information on the LO
- Lifecycle – history and versioning of the LO
- Meta-Metadata – information on the LO's metadata
- Technical – technical requirements of the LO
- Educational – educational and pedagogical information on the LO
- Rights – intellectual property rights of the LO
- Relation – relation to other LOs
- Annotation – comments on educational use of the LO
- Classification – description of the LO according to a particular classification system

Portal Connectivity. Taking over the decision from the cultural heritage objects portal, the learning objects portal also uses OAI-PMH as the protocol for harvesting metadata to and from external learning repositories.

Portal Components and Tools. The repository layer of the LO portal consists of two components. On the one hand this is the learning path template repository, holding structured path templates which represent particular pedagogical models, as well as open templates, which do not force the use of a certain structure but allow the user to freely design the components of their learning activities. On the other hand there is the learning path repository, storing the actual implemented learning paths.

On top of these repositories search services provide the features of searching for and retrieval of cultural heritage object contents, and secondly the searching for and retrieval of LOs like the structured and open learning paths and their templates. In both cases project-internal sources as well as external portals' contents can be discovered, like Europeana for the CHOs and the OSR portal for the LOs, and the internal portals' contents are also provided for consumption by these external portals.

Finally there will be a rich set of tools and user interfaces exploiting the portal's features, like a tool for editing the CHOs' metadata, which at the same time integrates certain sets of vocabularies according to the respective CHO's collection background, to ensure the use of common classifications for the metadata. Apart from that there will be web-based tools for the implementation of LP templates and the actual LPs, mainly to be used by museum and school educators and also incorporating the discovery of the CHO and LO contents to be used during the implementation. Finally web frontends provide the actual application of the pathways to the end users, like following the contained learning activities or using the pathway for navigation through the actual museum. A faceted search interface similar to the one developed in the MACE portal will allow the convenient search and navigation through the portals' underlying CHO and LO contents. Apart from that it is also intended that these tools and features will be used by interactive installations on 3D interfaces inside the NHMs which allow 'minority-report-like' to navigate through and interact with the contents. Figure 3 shows the basic layers of the LO repository with the respective components contained, their interaction and the target users.

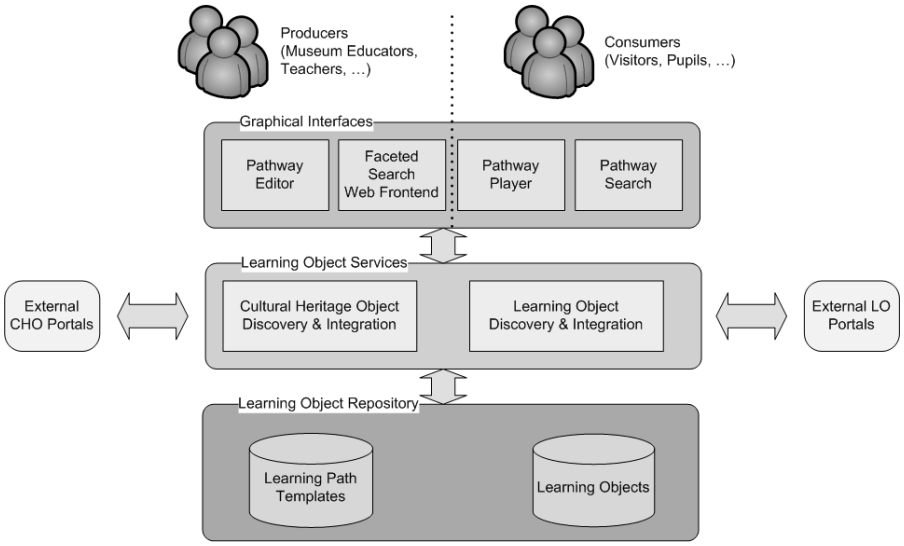

Fig. 3. The Natural Europe LO Repository

5 Conclusion

In this paper we presented the Natural Europe project and its targeted goal of exploiting the digital collection of cultural heritage items in natural history museums and connecting them to federated repositories for educational purposes. To achieve this, the architectural design is based on a continuous requirements engineering process whose outcomes, as we have shown, define the features of the Natural Europe system architecture, connecting a cultural heritage portal and a learning objects portal. Those are to be connected with other existing portals, CHO as well as LO, which also influenced the system architecture. The Natural Europe architectural design has by now reached a first completed status, the basic components of the CHO portal already being applied by the participating NHMs and the LO portal components being introduced and brought to use to the stakeholders. The next step will be the implementation of the interaction of the particular CHO and LO components with each other, followed by bi-directional content exchange with external portals. These steps will be accompanied by regular validations and performance tests, whose results affect the further implementation of the architecture and ensure the required functionality of the system.

1 References

1. Europeana: think culture, http://www.europeana.eu/
2. DELOS Network of Excellence on Digital Libraries, http://www.delos.info/
3. LOGOS: Knowledge-on-Demand for Ubiquitous Learning, http://www.logosproject.com/
4. e.KnowNet, http://www.e-knownet.eu/
5. MACE – Metadata for Architectural Contents in Europe, http://portal.mace-project.eu/
6. Natural Europe, http://www.natural-europe.eu/
7. Arctic Centre – University of Lapland, http://www.arcticcentre.org/
8. Hungarian Natural History Museum, http://www.nhmus.hu/
9. Jura Museum Eichstätt, http://jura-museum.de/
10. National Museum of Natural History – University of Lisbon, http://www.mnhn.ul.pt/
11. Natural History Museum of Crete – University of Crete, http://www.nhmc.uoc.gr/
12. Estonian Museum of Natural History, http://www.loodusmuuseum.ee/
13. Europeana Semantic Elements Specification, Version 3.4, http://www.version1.europeana.eu/c/document_library/get_file?uuid=77376831-67cf-4cff-a7a2-7718388eec1d&groupId=10128
14. Dublin Core Metadata Initiative: Dublin Core Metadata Element Set Version 1.1, http://dublincore.org/documents/dces/
15. Definition of the Europeana Data Model Elements, Version 5.2.2, http://www.version1.europeana.eu/c/document_library/get_file?uuid=7ef62859-4846-4697-9285-6bde99d976f5&groupId=10605
16. The Open Archives Initiative Protocol for Metadata Harvesting, Version 2.0, http://www.openarchives.org/OAI/openarchivesprotocol.html
17. IEEE 1484.12.1-2002 Draft Standard for Learning Object Metadata, http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf