Bringing Environmental Culture Content into the Europeana.eu Portal: The Natural Europe Digital Libraries Federation Infrastructure

Konstantinos Makris, Giannis Skevakis, Varvara Kalokyri,
Nektarios Gioldasis, Fotis G. Kazasis, and Stavros Christodoulakis

TUC-MUSIC, Lab. Of Distributed Multimedia Information Systems,
Technical University of Crete, University Campus, 73100, Chania, Greece
{makris,skevakis,vkalokyri,nektarios,
fotis,stavros}@ced.tuc.gr

Abstract. The aim of the Natural Europe project [1] is to improve the availability and relevance of environmental culture content for education and life-long learning use, in a multilingual and multicultural context. Cultural heritage content related with natural history, natural sciences, and nature/ environment preservation, is collected from six Natural History Museums (NHMs) around Europe into a federation of European Natural History Digital Libraries that is directly connected with Europeana.eu. We present here the Natural History Digital Libraries Federation infrastructure along with the appropriate tools and services that (a) allow the participating NHMs to uniformly describe and semantically annotate their content according to international standards and specifications, (b) interconnect their digital libraries, and (c) expose metadata records for Natural History cultural heritage objects to Europeana.eu.

Keywords: Europeana, cultural federation, digital curation, preservation metadata.

1   Introduction

In an era where natural history and environmental education inadequacy in formal and informal contexts is becoming an increasingly challenging issue, harvesting the potential of European digital libraries appears as a very attractive option. Natural History Museums (NHMs for short) can play an important role to the reconstruction of our environmental culture since they are "key tools" of social and cultural development and can promote the harmony between humanity and nature. NHMs are generally well prepared to meet the above challenge. Most of them are quite innovative institutions dealing at the same time with conservation of collections, scientific research, and cultural and educational activities. Consequently, they have the possibility to conceive their own special ways of mediation and they have the opportunity to deal with environment as a natural part of our general culture.

However, an impressive abundance of high quality digital content that is available in NHMs around Europe remains largely unexploited due to a number of barriers, such as: the lack of interconnection and interoperability between the management

systems of NHMs, the lack of centralised access through a European point of reference like Europeana, as well as the inadequacy of current content organization and the metadata used.

The Natural Europe project offers a coordinated solution at European level that aims to overcome the aforementioned barriers in order to amplify the exploitation of natural history heritage into formal and informal learning processes. Among the main objectives of the Natural Europe project is the development of a Natural History Digital Libraries Infrastructure which offers appropriate tools and services that (a) allow the participating NHMs to uniformly describe and semantically annotate their content according to international standards and specifications, (b) interconnect their digital libraries, (c) expose their metadata records for Natural History (NH for short) Cultural Heritage Objects (CHO for short) to Europeana.eu.

The rest of this paper is organised as follows. Sections 2, 3, and 4 present the infrastructure developed to support the aggregation of Environmental Heritage resources (spread in European Natural History Museums) and their further dissemination to external organizations. Section 5 presents an indicative information flow in our case study. Section 6 concludes and presents the work that is planned to take place in the next future.

2   The NH Digital Libraries Federation Infrastructure

The NH Digital Libraries Federation Infrastructure aims to aggregate natural history content metadata from NHMs and Europeana.eu in order to exploit them for learning purposes. Moreover, it aims to expose the aggregated metadata records to Europeana.eu. To achieve these objectives a set of tools and services have been developed and deployed in order to ensure that possibly diverse legacy natural history metadata will be unified, semantically enriched, aggregated, preserved, and finally harvested by Europeana.eu. The federation's stakeholders and their basic requirements are:

1. **The Natural History Museums.** The participating NHMs are the main content providers whose metadata are aggregated into the federation. With respect to NHMs the federation architecture ensures that at least the following requirements are met: a) *Legacy Metadata Migration*. Any existing metadata is properly migrated (mapped, imported, and handled) to the environment; b) *Metadata Unification.* All the contributed metadata records are encoded following the ESE [2] and EDM [3] schemas; c) *CHO Publishing*. Each contributed CHO or a representation needs first to be published and web accessible (in its original quality or not). d) *Semantic Multimedia Annotation.* NHMs have the ability to classify and/or annotate/link their contributed CHOs with terms from common (widely-accepted) controlled vocabularies (with concepts, places, persons, etc.) and/or taxonomies; e) *Standardized Metadata Dissemination.* The metadata records of the contributed CHOs are exposed through standard protocols, schemas, and formats.
2. **The Europeana.eu Cultural Heritage Portal.** The NH metadata that the Natural Europe aggregates will be made available through the Europeana.eu Portal. On the other side, the federation also harvests from it NH related metadata that is already available there. Currently the Europeana.eu metadata dissemination process is

based on web service and widget technologies that allow to search and display the Europeana collections in external applications[1]. For the metadata submission process, the following requirements have to be met: a) *Conformance to ESE/EDM.* Currently, the Europeana's metadata submission process requires that providers should encode their metadata records according to the Europeana Semantic Elements specification (ESE for short), while in the near future it will be able to accept also metadata that conform to the Europeana Data Model (EDM for short) specification; b) *OAI-PMH Compliant Metadata Dissemination.* In order to submit the aggregated metadata to Europeana.eu, the federation offers appropriate OAI-PMH [6] service interfaces; c) *CHO Persistent Identification.* Each contributed CHO is persistently identified and addressed so that local (to museum) identification changes do not affect Europeana's resolving process.

3. **Third-Party Federations.** Integration with third-party federations (e.g. BHL [4], Organic.Edunet [5], etc.) imposes specific requirements. For that the NH Digital Libraries Federation infrastructure offers an OAI-PMH [6] based metadata dissemination service.
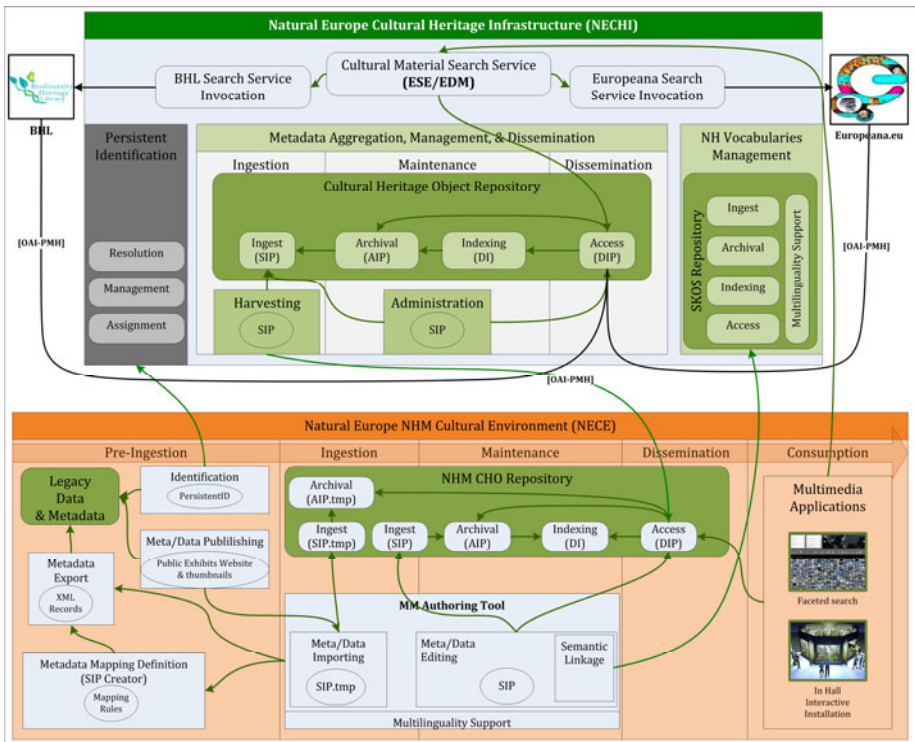

Fig. 1. The Natural History Digital Libraries Federation Architecture

---

[1] Europeana also aims to provide an OAI-PMH target service that will enable external applications to harvest the Europeana collections.

Fig. 1 illustrates the Natural History Digital Libraries Federation Architecture consisting of two nodes: a) The Cultural Heritage Infrastructure (the Federal) Node where the metadata aggregation, management, and dissemination is taking place. Moreover, this infrastructure provides NHMs with appropriate services for semantic annotation and persistent identification of their CHOs; and b) The NHM Cultural Environment (the Federated) Node that provides appropriate CHO curation and management tools and services, while also defines specific practices and processes in order to achieve the requirements discussed above.

3   The Natural Europe Cultural Heritage Infrastructure

The Cultural Heritage Infrastructure (NECHI for short) accommodates software components that can be categorized under the following three high level functions:

1. Metadata Aggregation, Management, and Dissemination of CHOs contributed by the participating NHMs.
2. Shared Knowledge Management: Management of shared vocabularies as well as brokerage of other vocabulary or taxonomy accessing services.
3. Persistent Identification and Resolution: Services that allow local identification modifications to remain transparent.

3.1   Metadata Aggregation, Management and Dissemination

Metadata Aggregation, Management, and Dissemination is the core theme of the NECHI. This functionality actually refers to a complete information preservation life-cycle as defined in the OAIS Reference Model [7], where the information producers are the participating NHMs and information consumers are the Europeana.eu Portal, and other third-party federations like the BHL and Organic.Edunet Portals.

In the NECHI the OAIS reference model is materialised by the Cultural Heritage Object Repository (CHO Repository for short) that accommodates appropriate software components that span the Ingestion, Maintenance, and Dissemination phases for the CHO metadata contributed by NHMs. In each phase, appropriate information packages are manipulated by the repository.

Information Packaging. A clear definition of information is central to the ability of an OAIS to preserve it and is one of the first steps in designing such a system. Information is manipulated by an OAIS in the form of Information Packages. In Natural Europe an information package may accommodate both content information (CHOs) as well as descriptive information (CHO metadata and administrative information).

Every submission of information to an OAIS by a Producer, and every dissemination of information to a Consumer, occurs as one or more discrete transmissions of Information Packages:

- The Submission Information Package (SIP for short) refers to the information package that is submitted to the system by a Producer in the Ingestion Phase. A SIP for the CHO Repository is created as a result of a metadata harvesting process or as a result of a metadata modification process from the CHO

Metadata Administration Interface. In the first version of the CHO repository, SIPs conform to the ESE Specification Version 3.4. However in future releases, SIPs for this repository will be based on the EDM specification.

- The Archival Information Package (AIP for short) refers to the information packages that are archived by the CHO Repository in the Maintenance Phase. Within the CHO Repository one or more SIPs are transformed into one or more AIPs for storage according to its internal storage specification (e.g. collection-based, record based, aggregation-based, etc.). Each AIP has a complete set of administration information for the associated descriptive information. AIPs will be also indexed on various descriptive information attributes in order to be efficiently accessed and disseminated.

- The Dissemination Information Package (DIP for short) refers to the information package that the CHO Repository provides in response to a request for accessing all or a part of an AIP from a Consumer in the dissemination phase. A DIP may include part of an AIP (e.g. as a result of a search request), a complete AIP (as a result of an access request), or collections of AIPs (e.g. as a result of a harvesting request), and it may or may not contain complete descriptive information.

Software Modules & Services. As illustrated in Fig. 1, the Metadata Aggregation, Maintenance and Dissemination functionality is implemented by a) the CHO Repository, b) the Harvesting Module, and c) the Administration Module.

The CHO Repository is responsible for the maintenance (storage, management) and dissemination of the aggregated CHO Metadata. It accommodates modules for ingestion, archival, indexing, and accessing of CHO Metadata.

The Harvesting module is responsible for the scheduled harvesting of CHO metadata from the participating NHMs. For that, this module acts as an OAI-PMH Harvester application that is issuing specific harvest requests to NHMs' metadata repositories following a predefined schedule.

The Administration module allows administrators to manage the CHO Repository AIPs and their administration information. It also allows them to manage user accounts and privileges, as well as to schedule harvests from specific NHMs. This module allows also content providers to inspect (preview) and manage (accept, reject) harvested metadata collections.

The Cultural Material Search Service exploits functionality provided by the Access module (and backed up by the Indexing module) in order to search for CHO descriptions (in the form of ESE/EDM) aggregated by the participating NHMs and stored in the CHO Repository. Moreover, it is routing search queries to other federations or portals like Europeana.eu and BHL, in order to allow the exploitation of Natural History related content (the metadata of which are stored in such federations) for other purposes. The results of this service are returned as DIPs in the form of ESE/EDM descriptions or references to them.

3.2   NH Shared Knowledge Management and Dissemination

In the context of Natural Europe the participating NHMs enrich the description of their contributed CHOs through a semantic annotation process. In order for this semantic annotation process to be effective in terms of exploitation (i.e. others can

understand the annotations), a community-wide accepted taxonomy [8] has been adopted. Mappings between the classifications that the NHMs use and the standard taxonomy have been defined.

To support this, NECHI accommodates a specific sub-system for managing these taxonomies and/or classification schemes (vocabularies) as well as allowing museum curators to search and locate specific entries into them. Moreover, this sub-system may also act as a proxy for accessing vocabularies from other NH related communities like BHL, uBio [9], etc. if such a requirement is considered important for the relevant stakeholders (e.g. NHMs). In the future when EDM will be widely adopted, the Shared Knowledge Management and Dissemination sub-system will offer access to shared vocabularies and/or authority files officially adopted by the Europeana Semantic Layer (ESL for short) [10] in order to allow a fully-fledged EDM usage.

Natural Europe shared vocabularies have been represented as RDF resources based on the SKOS model [11]. Thus, this sub-system offers functionality for publishing (ingest), storing (archival), indexing, and searching SKOS vocabularies.

The adoption of the SKOS model for the representation of the shared vocabularies is motivated by the fact that SKOS has been also adopted for the development of the ESL.

3.3   Persistent Identification and Resolution

Persistent Identification of material is a well identified requirement in many application domains and becomes more vital as we move to the web of data. Europeana.eu has already identified [12][13] this requirement and has put persistent identification of aggregated content in top priority.

Persistent Identification is very important as it enables digital objects to have a stable access point over the Internet. Persistent identifiers (PIDs for short) are resolved by appropriate resolution services that resolve them to the current valid address of a digital object.

Following that, Natural Europe has taken care so that the aggregated CHOs are persistently identified. For that, NECHI accommodates a specific sub-system that allows museums to define PIDs for their contributed CHOs (if they don't already do so) using the Purl system [14]. This sub-system provides a specific identification namespace for each participating NHM, under which the NHM persistently identifies its contributed CHOs. Moreover, this sub-system provides a resolution service which is be able to resolve a PID to the registered CHO address.

This approach is aligned with the specification of the Europeana Resolution and Discovery Services (ERDS for short) [15] which keeps track of namespaces of persistent identifiers and of the institutions that are responsible for those namespaces. With the help of this information, the ERDS can forward resolving requests to the right local resolver and present the returned link.

4   The NHM Cultural Environment

The NHM Cultural Environment (NECE for short) node provides: a) the specification of the Metadata Management Life-cycle process along with the tools and services needed in order to ensure that existing collections and their descriptions will be migrated, annotated, preserved, and properly disseminated (mainly to NECHI),

and b) the advanced Cultural Heritage Multimedia Consumption Applications that enable NHMs to present (acting as Information Consumers) Natural History content to their visitors (physical or virtual ones); content not only from their collections, but also from other collections reached through the NECHI.

4.1 The NHM Metadata Management Life-Cycle Process

In Natural Europe the participating NHMs contribute (metadata about) a large number of Natural History related CHOs which are aggregated by the project and exploited for educational purposes. In order to ensure that a) any existing metadata will be exploitable in this context and b) the new knowledge will be built on the same basis for all participating NHMs, a metadata unification task is required. Moreover, these unified metadata need to be preserved in a dedicated metadata repository able to support the enrichment activities over them. For supporting these requirements, NECE provides the NHM CHO Repository which is able to handle metadata in the form needed by Natural Europe (ESE/EDM). This repository along with the MM Authoring Tool[2] which has been implemented on top of it, are the cornerstone components of the NECE at each NHM side as far as the Metadata management and dissemination is concerned. CHO metadata is ingested to, preserved at, and disseminated from, this repository.

The complete life-cycle process that the NECE defines for NHM (contributed) metadata management comprises four phases: a) pre-ingestion phase, b) ingestion phase, c) maintenance phase, and d) dissemination phase.

Except for the preparatory pre-ingestion phase, the other three phases (ingestion, maintenance, dissemination) are standard phases defined for an Open Archival Information System. In these phases each NHM uses as a front-end Application (the MultiMedia (MM) Authoring Tool) for the semantic annotation of its CHOs, and as a back-end infrastructure the NHM CHO Repository that offers standard OAIS functionality for the enriched metadata. In the sub-sections below we discuss the various phases of the metadata management life-cycle process along with the tools, repositories, and services of the NECE that NHMs will be using in each phase.

Pre-Ingestion Phase. In the pre-ingestion phase, each NHM ensures that it's existing CHO collections and their metadata descriptions will be appropriately migrated into Natural Europe. This includes:

1. Meta/Data Publishing of contributed CHOs: Unless already did so, each NHM can currently use the provided MM Authoring Tool in order to publish each contributed CHO to the web, so that it will be accessible in one click from end users (users of Europeana.eu, Natural Europe, BHL, etc.). The tool automatically publishes the CHO and creates a thumbnail for the contributed CHO.
2. Identification of contributed CHOs: Unless already did so, each NHM will be able in the upcoming release of the MM Authoring Tool to persistently identify the CHOs that will contribute to the project using appropriate services provided by the Persistent Identification sub-system of NECHI. As mentioned, each NHM is provided with a specific identification namespace under which it will permanently identify its contributed (at least) CHOs.

---

[2] The running version of this tool is available at `http://147.27.41.103:8080/mmat`

3. Metadata Unification: Metadata unification (as discussed above), is implemented during the pre-ingestion phase in two steps (following the Europeana.eu content provision practices); firstly, the legacy NHM metadata for the contributed CHOs is provided as record based XML documents and secondly, NHM metadata experts, along with Natural Europe technical experts map these legacy metadata to the ESE specification. Metadata Mapping Definition is done using the Europeana.eu SIP Creator application [16]. SIP Creator is a web-launched GUI tool that allows any record based XML schema (inspected from XML data) to be mapped to ESE schema. The output of this process is a set of mapping rules that are used in the ingestion phase from the MM Authoring Tool for transforming the legacy metadata into ESE format and importing them into NHM CHO Repository.

Ingestion Phase. The ingestion phase refers to both importing existing metadata (after applying the mapping rules) to the Natural Europe environment and enriching metadata through a semantic annotation process.

1. Regarding the importing of existing metadata, the MM Authoring Tool provides the functionality to load existing collection metadata, load the appropriate mapping rules, apply these rules to the loaded metadata producing a Submission Information Package (for the underlying repository) that conforms to ESE specification. This
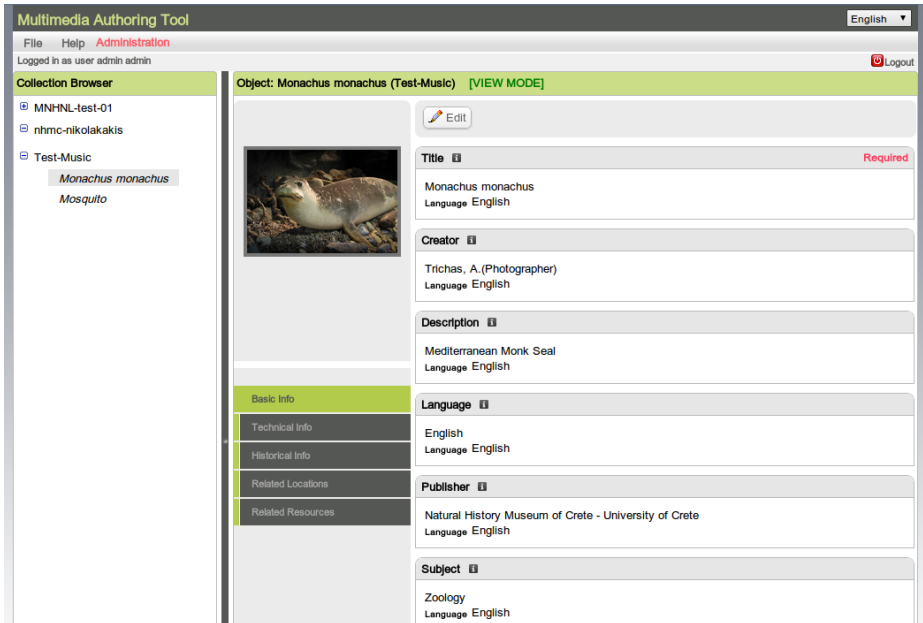

Fig. 2. The MultiMedia Authoring Tool in use

SIP, is stored into the NHM CHO Repository with a pending status (i.e. as a temporary AIP). Museum curators have then the ability to inspect the imported collection metadata, modify, reject, or accept it. Once accepted, the imported collection is permanently stored and properly indexed into the NHM CHO Repository as a normal AIP. In this importing phase, the MM Authoring tool allows museum curators to define the identification scheme for the imported records, taking into account any existing identification policy. An indicative screen-shot of this tool is illustrated in Fig. 2.

2. As far as ingestion through the normal metadata curation/annotation activity is concerned, the MM Authoring tool allows museum curators to maintain (create/view/modify/enrich) CHO metadata. This activity is facilitated by access and concurrency control mechanisms provided to ensure security, integrity, and consistency of contents.

Maintenance Phase. The maintenance phase refers to storage and management of CHO metadata. At each museum, a complete metadata management solution is provided consisting of: a) the NHM CHO Repository as a back-end infrastructure, and b) the MM Authoring Tool as a front-end Metadata Authoring Tool. The NHM CHO Repository is implementing the OAIS Reference Model with all its discrete entities. For the Maintenance phase, the Archival, and Indexing modules provide the required functionality. As mentioned, the stored AIPs are currently based on the ESE specification, while in future releases they will be based on the EDM specification.

On the other hand, maintenance covers also the day-to-day metadata curation activities of a museum. For that, the MM Authoring tool provides museum curators with a intuitive Graphical User Interface that allows them to easily a) view, edit, delete and create CHO metadata; b) create and manage CHO collections; and c) semantically annotate (interlink) CHOs using shared knowledge systems (SKOS vocabularies and taxonomies, or other RDF resources). The MM Authoring tool currently supporting the ESE specification (i.e. it is producing and manipulating Information Packages conforming to ESE); however, in future releases, it will be able to work with EDM allowing NHMs to exploit the full potential of semantic web and interlink their CHOs with other resources on the web of data (e.g. link them with ESL resources, other published CHOs, authority files for places, events, persons, etc.).

Finally, the tool is multilingual and has been provided to the NHM partners' languages. Moreover, it allows the definition of metadata in any language. Elaborating more on this, multilingual CHO descriptions are provided using the XML "Lang" attribute although Europeana.eu does not currently directly exploits it.

5   Indicative Flow of Information

As a pilot case we consider a digital representation of an exhibit of the Natural History Museum of Crete [17] along with its descriptive metadata. Below (Fig. 3) we provide its description as defined in the ESE specification that is currently supported by Europeana.eu as created by the MM Authoring Tool. The same description will be provided in EDM (Fig. 4), when fully supporting it, and interlinked with other resources of the semantic web.

Node	Content
▽ ⓔ metadata	
ⓐ xmlns:dc	http://purl.org/dc/elements/1.1/
ⓐ xmlns:europeana	http://www.europeana.eu
ⓐ xmlns:dcterms	http://purl.org/dc/terms/
▽ ⓔ record	
ⓔ dc:coverage	Greece
ⓔ dc:creator	Trichas,A.
ⓔ dcterms:issued	09/11/2010
ⓔ dc:description	Mediterranean monk seal (common name). NHMC exhibit.
ⓔ dc:format	TIFF
ⓔ dc:identifier	35651
ⓔ dc:language	English
ⓔ dc:publisher	Natural History Museum of Crete-University of Crete
ⓔ dc:rights	NHMC
ⓔ dc:source	Natural History Museum of Crete-University of Crete
ⓔ dc:subject	Zoology
ⓔ dc:subject	MAMMAL
ⓔ dc:title	Monk seal Monachus monachus
ⓔ dc:type	Image
ⓔ dcterms:alternative	Monachus monachus
ⓔ dcterms:created	12/10/2002
ⓔ dcterms:spatial	NHMC, Knossou Avenue, Irakleiou, Kriti
ⓔ europeana:collectionName	NHMC data_set_1
ⓔ europeana:country	greece
ⓔ europeana:dataProvider	Natural History Museum of Crete-University of Crete
ⓔ europeana:isShownBy	http://www.nhmc.uoc.gr/zoology/35651.html
ⓔ europeana:language	el
ⓔ europeana:object	http://www.nhmc.uoc.gr/zoology/35651_tn.jpg
ⓔ europeana:provider	Natural History Museum of Crete-University of Crete
ⓔ europeana:type	IMAGE
ⓔ europeana:uri	http://www.europeana.eu/resolve/record/nhmc_dataset_1/568A74590C376DEF30D8E9CD6C23671902245EA6

Fig. 3. Example Description of Natural History Museum Object (ESE)
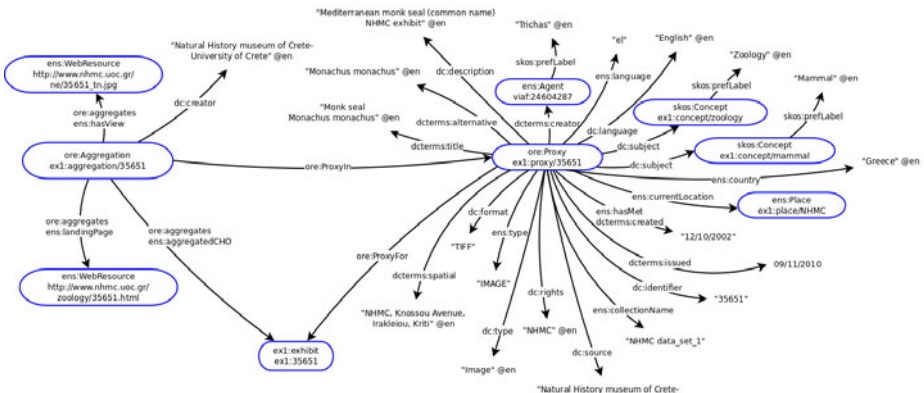

Fig. 4. Example Description of Natural History Museum Object (EDM)

6 Conclusions

In this paper we presented the infrastructure developed in the context of the Natural Europe project to support the aggregation of Environmental Culture resources (spread in European Natural History Museums), their exploitation for education, and their

further dissemination to external organizations. The Natural Europe federation infrastructure consists of: a) The federated node that accommodates tools and services to support the complete metadata (and optionally data) life-cycle management process (ingestion, preservation, dissemination). Apart from the deployed tools and services on the federated nodes, the project has set up specific work-flow processes for the publication, migration (from existing systems), and semantic description, of contributed resources. b) The federal node accommodating services and applications to support the aggregation, exploitation, and further dissemination of resource metadata using standard OAI-PMH services. The federal node provides also services for shared knowledge management (SKOS vocabularies and authority control files), and persistent identification (PURL maintenance and resolving).

Currently the federated node has been provided for use to six European NHMs with a positive feedback regarding the usability and functionality of the tools and the provided functionality. It is expected that the federal node, which is currently under development, will begin to harvest the contributed CHOs within the next few months. According to the Natural Europe project's work-plan, the contributed CHOs will be exposed to Europeana.eu portal by the end of this year.

References

1. The Natural Europe Project, `http://www.natural-europe.eu`
2. Europeana Semantic Elements Specification V.3.4,
   `http://version1.europeana.eu/c/document_library/get_file?uuid=77376831-67cf-4cff-a7a2-7718388eec1d&groupId=10128`
3. Europeana Data Model Definition V.5.2,
   `http://version1.europeana.eu/c/document_library/get_file?uuid=aff89c92-b6ff-4373-a279-fc47b9af3af2&groupId=10605`
4. Biodiversity Heritage Library (BHL),
   `http://www.biodiversityheritagelibrary.org/`
5. Organic.Edunet Portal. Learning Material on Organic Agriculture in Europe,
   `http://portal.organic-edunet.eu/`
6. Open Archives Initiative – Protocol for Metadata Harvesting V.2,
   `http://www.openarchives.org/OAI/openarchivesprotocol.html`
7. ISO 14721:2003 Open Archival Information System (OAIS) Reference Model,
   `http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683`
8. The Catalogue of Life, `http://www.catalogueoflife.org`
9. Universal Biological Indexer and Organizer (uBio), `http://www.ubio.org`
10. The Europeana Semantic Data Layer,
    `http://www.europeanaconnect.eu/documents/D1.1.2_eConnect_Semantic_Layer.v1.0_UVA.pdf`
11. W3C: Simple Knowledge Organisation System (SKOS),
    `http://www.w3.org/TR/2009/REC-skos-reference-20090818/`

12. Cousins, J.: Europeana: Access to Culture for All?,
    `http://www.conference.bvoe.at/presentations/cousins.pdf`
13. The Europeana Strategic Plan 2011-15:
    `http://version1.europeana.eu/c/document_library/get_file?uui`
    `d=c4f19464-7504-44db-ac1e-3ddb78c922d7&groupId=10602`
14. Persistent Uniform Resource Locators (PURL), `http://www.purl.org`
15. EuropeanaConnect: D5.4.1Europeana Resolution Service,
    `http://www.europeanaconnect.eu/documents/D5.4.1_eConnect_Eur`
    `opeanaResolution_Service_v1.0_DNB.pdf`
16. Europeana Labs: SIP Creator, `http://europeanalabs.eu/sip-creator/`
17. The Natural History Museum of Crete (NHMC), `http://www.nhmc.uoc.gr`