ECK Guide

Sharing collections with the Easy Connection Kit

ECK GUIDE

Sharing collections with the

Easy Connection Kit

ECK Guide
Sharing collections with the Easy Connection Kit

Texts by:
Gábor Palkó, Petőfi Literary Museum
Laurent Meese, Royal Belgian Institute of Natural Sciences
Marco Streefkerk, Digital Heritage Netherlands
Neil Smith, Knowledge Integration
Inés Matres, Institute for Museum Research, Preussischer Kulturbesitz

Edition and Design: Inés Matres, Institute for Museum Research, Preussischer Kulturbesitz

Reviewed by Laura Miles, Collections Trust

Pictures courtesy of:
Ecomuseu Municipio do Seixal
Ethnologisches Museum – Staatliche Museen zu Berlin

Berlin, February 2015.

CONTENTS

INTRODUCTION

In 2012 the Europeana Inside project started with the vision that cultural institutions should be able to easily provide content to Europeana. With this vision and the goal of lowering barriers to content sharing, the ECK – Easy Connection Kit – has been created. The ECK is a piece of software and documentation toolkit that has been integrated into existing commercial and open source Collection Management Systems as well as into aggregators. Moreover, the components that make up the ECK can be integrated into other systems for free.

In this document we present a guide to the ECK. Cultural heritage institutions, software vendors or anyone setting up a new aggregation service who want to know more about the opportunities the ECK can offer will find information here. Furthermore this document explains how the ECK came about and how it has had an impact on the work of institutions who are sharing information about collections with the ECK. Furthermore, with this guide we will introduce some simple steps to manage the sharing of collections with others and a workflow that can be integrated into the documentation process of any given collecting institution.

ECK-ready systems contain a high quality, resilient service that enables cultural heritage institutions to manage their participation in Europeana and similar data-sharing initiatives within their existing systems and respecting their processes.

THE STORY OF THE ECK

Providing access to collections and distributing information about them are activities anchored in the mission statements of any cultural heritage institution. Sharing information about cultural objects can be done through multiple channels. Most memory institutions relate this task with exhibitions, catalogues and the presentation of highlights on their own website. However, even though digital technologies have opened up the possibilities and diversified the options in which objects can be presented to the public, often technical and legal aspects remain a challenge.

Memory institutions, such as museums and libraries, have a long tradition in the field of documentation. Providing a historical context and information on provenance for each of the objects in their collections is a core task for cultural heritage professionals. As a result of this work, the collection catalogue, where all object related data is gathered, can only be described as an invaluable source of information about heritage.

Collection Management Systems are software products developed by commercial or non-profit organisations acquired by cultural institutions and installed in their local systems. Each institution, depending on its capacity, will choose a different product or company to manage their digital collections; but all systems are subject to continuous upgrade and change. Often, newly acquired systems are adapted to the institutions' needs before even putting them to use.

Europeana Inside departed from these facts and examined what steps are needed in order for the carefully gathered and documented information, to be selected, prepared and shared with cultural platforms such as Europeana.

The goal that Europeana Inside set was to level the playing field and lower barriers any participation barriers that institutions encountered when sharing. This was done under

the premise that this should be achieved respecting processes used by heritage professionals and keeping the software they are accustomed to.

A group of companies with experience in the development of documentation systems for cultural heritage institutions joined forces and worked in close collaboration with their clients (museums and libraries using their products) to create an innovative solution that allows institutions to face the challenge of sharing collections. The Easy Connection Kit (ECK) was implemented in multiple products on the same premises:

- ► Invisible and integrated into the system that the user is already used to;
- ► Robust, through thorough testing of all software components;
- ► Flexible in its design, with no fixed workflow but adopting a workflow that is appropriate to the needs of its users;
- ► Modular, service oriented architecture (SOA) with RESTful APIs between components;
- ► Using a common intermediate metadata model (LIDO) in order to ensure that rich information is available for further re-use, in multiple end-points;
- ► Configurable, avoiding hard coding of functions and allowing flexibility by applying different configuration parameters.

By helping to develop the ECK, a number of museums and libraries all over Europe have already contributed hundreds of thousands of records to Europeana with an ECK-ready system. In the following section of this guide cultural institutions have provided an account of their experiences and we analyse how an aggregator could also benefit from installing ECK components in order to connect directly with the collections' systems of its providers.

BENEFITS FROM USING ECK

A MUSEUM

"Adapting ECK allows us to manage records that appear in Europeana from our system "

Petőfi Literary Museum, Budapest

The main motivation that drove the Petőfi Literary Museum[1] to participate in Europeana Inside was to adapt the ECK (Easy Connection Kit) to their own needs in order to manage the records that appear on Europeana. Two and a half years later Gábor Palkó, research director of the museum reviews his experience in the project.

The Petőfi Literary Museum engages with audiences mainly through publishing the collection in their own OPAC and by organising exhibitions. Additionally, the museum engages with initiatives to spread their content and give access to their objects through external online platforms. In the past the museum has participated in the Athena project (2008-

1 www.pim.hu

2011)[2], acting as a liaison for engaging institutions to participate in Europeana (e.g. the Hungarian Open air Museum, the Hungarian National Gallery and the Museum of Fine Arts Budapest). During Athena, they gave access to 1194 metadata records through Europeana. They have also participated in the Google Art project[3], where digitalized images of 48 works of art where uploaded to Google´s virtual museum through the Google Online Content Upload System. In the on-going Athena Plus project[4], the museum will be giving access to 5000 digital objects and will be sending their metadata to Europeana.

By participating in Europeana Inside, the museum has given access to a further 10, 000 records which showcases highlights of their collections.

The workflow below illustrates the workflow the Petőfi Literary Museum followed to share their content in the framework of Europeana Inside.

1. Selection of the records (the collection) which included:

  ► The book collection which had no copyright attached to it. Each record is accompanied with biographical information on the author which distinguishes these as museum objects rather than from libraries;

  ► The museum also collects book covers, something that libraries sometimes do not document or preserve properly;

  ► Other objects in the collections are photographs, some of which are artistic photographs made by authors or in relation with the authors; others are

---

2 www.athenaeurope.eu

3 www.google.com/culturalinstitute/collection/museum-of-literature-petofi

4 www.athenaplus.eu

authors' portraits;

- ► One collection that was also provided to Europeana was the relic's collection which holds personal objects belonging to national literary figures;

- ► The last collection was the Fine Arts collection of illustrations, portraits and drawings on authors or made by authors.

2. Checking the completeness and accuracy of data and that all metadata records have been assigned with at least a digital object (a picture), this is provided in a routine that runs in the background in their database system (Qulto[5]).

3. Mapping. To map elements is tricky and requires knowledge of the context of the objects that make up a collection. A 'boot' can be an object in the clothing section of a collection or it can be an illustration of a boot in a painting. There is no automatic process that can provide content recognition; therefore the curator needs to be involved in this process.

4. Upload or push to the dark aggregator[6]. This is where validation against LIDO / EDM happens.

5. The Petőfi Literary Museum can today manage enrichments[7] with their collection management system. These enrichments can improve the technical information on subject headings used in their catalogue process, as well as several other improvements.

---

5 http://qulto.eu/

6 The Dark Aggregator was set up for developing the ECK. It contains all functionalities and acted during Europeana Inside as data repository between providers and Europeana.

7 To know more about enrichment process in Europeana visit: http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+force+multilingual+semantic+Enrichment

To know more on the re-ingestion capacity read on the section "Workflow for sharing collections"

BENEFITS FROM USING ECK

A LIBRARY

"We have obtained the knowhow and software to share collections in a more transparent way"

Royal Belgian Institute of Natural Sciences, Brussels

By participating in this best practice network the Royal Belgian Institute for Natural Sciences has gained access to technical knowledge and expertise from other European heritage organisations and technical and commercial partners. Laurent Meese, curator at the Institute has explained how.

The Royal Belgian Institute of Natural Sciences houses a precious collection of zoological, anthropological, paleontological, mineralogical and geological materials and data. They use the DaRWIN[8] database as a main gate to glimpse the extent and diversity of the collections. The DaRWIN database manages information on about 350,000 specimens stored in the institute's depositories.

---

8 https://darwin.naturalsciences.be/

In addition, the central library of The Royal Belgian Institute of Natural Sciences has 600,000 bibliographic records relating to natural sciences in general. Our bibliographic records are managed with several databases and are mainly hosted by their technical provider LIBIS[9]. These collections are held in two other bibliographic utilities: Antilope (catalogue for scientific serials) and Belgium.be (central catalogue for federal libraries).

The Royal Belgian Institute for Natural Sciences has been a content provider for The Biodiversity Heritage Library for Europe[10]. This is a project engaging 28 natural history museums, botanical gardens and other cooperating institutions. BHL-Europe provides a multilingual access point for biodiversity content with specific biological functionalities for search and retrieval through the Europeana.eu portal. We delivered the digitised materials to this project via the dark aggregator developed for Europeana Inside which is then connected to the organisations records with no need of a mediator.

Using ECK-functionalities, we now have a local tool to easily select and manage data for export. There is a better structured workflow, staff members understand the importance and know how to use the ECK.

Europeana Inside has raised awareness of the importance of providing correct copyright information and other types of Intellectual Property Rights. The local ECK helps this process by providing accurate copyright information to our data. In order to avoid legal limits, we carefully selected our materials and supplied all in-copyright and free domain content.

---

9 www.libis.be

10 www.bhl-europe.eu

The last improvement that ECK has brought to our local system is the capacity to return enrichments from Europeana. The addition of geonames or multilingual data is useful for us. We can then incorporate the information in our system. The main advantage of enriched content is that it provides more search options for cross domain research in a multilingual context.

BENEFITS FROM USING ECK

AN AGGREGATOR

"We provide a Network, tools and support"

Beeld en Geluid (Sound and Image), Hilversum

There are many reasons why a cultural institution would want to share information about their collections. This goal is rapidly becoming more anchored in the mission statements of collecting institutions and often is a requirement for government or external funding.

Aggregators can be good partners in achieving this goal.

The choice for a certain aggregator can depend on the domain, country or region of an institution. There are national, regional and thematic aggregators. National and regional aggregators are usually funded by governments, can offer support, help institutions solve technical issues and in the best cases also have specialized departments for the main cultural domains: libraries, museums or archives. Additionally, thematic aggregators can create a lasting network of professionals and experts in one field, thus pulling together efforts and sharing resources. There are also European projects that act as temporary aggregators.

The national strategies for digital heritage in the Netherlands put the national domain aggregators as points of exchange between heritage organisations in the country and Europeana. Beeld en Geluid runs the national cultural heritage aggregator: Digitale Collectie[11] that serves as a coordinator. As an aggregator, Digitale Collectie offers tools and support.

Digitale Collectie stimulates memory organisations to supply their collections using standards and shared technology. CMS vendors are considered an important partner for this process. The ECK and the functional requirements on which it was built provide useful support to connect heritage organisations to the aggregation network via their CMS. The support of the ECK allows aggregation in The Netherlands to become a more standardised, automated process lessening the workload at the aggregator side.

---

11 www.digitalecollectie.nl

WORKFLOW FOR SHARING COLLECTIONS

The ECK is a suite of open and free software components that can be implemented by Collection Management System vendors or aggregators. Since CMS and aggregators are developed to the needs of their users there is no fixed way in which the components have to act or interact. Instead the modules can be configured to fit into the existing systems and their workflows.

In order to show what kind of services the ECK provides, this section offers a possible workflow drawn from various implementations that can be integrated into the documentation process of any cultural heritage institution.

Step 1

MANAGING DATA

Managing collections is a core mission of any heritage organisation and sharing digital collections brings some additional requirements and tasks. The capacity of keeping track of which data is shared and with whom, as well as registering changes of data shared is a key function in ECK that helps maintain and synchronise data at different platforms.

Step 2

SELECTING DATA TO SHARE

Unless a cultural institution wants to share the whole collection information with one aggregator, parts of the collections have to be selected in accordance with the objectives and needs set by the aggregator (e.g. a collection of digital objects dealing with fashion design for a thematic aggregator such as Europeana Fashion[12]).

Selection, again, is a task for the collection curator. This process involves choosing which part of the collection can be shared, but it also means that certain pieces of information about each object (fields) can be selected. With most Collection Management Systems this step is easily done.

---

12 www.europeanafashion.eu

PREPARING DATA TO BE EXPORTED

Whichever aggregator service is chosen, metadata should be transformed to fit the target. Assuming this is Europeana, usually more than one transformation is needed: from the internal data model to the model the first aggregator is using and, eventually, to the data model of Europeana – EDM (Europeana Data Model)[13]. Each transformation that the metadata undergoes requires rules that define how the outgoing information has to be interpreted by the receiver. This process is called mapping. A successful mapping ensures that descriptions fit the institution's own objectives as well as the aggregator's minimal requirements.

The ECK includes a standard mapping via LIDO[14] which is the most popular exchange format for museums. There is also a mapping available for MARC[15], which is a widely used metadata standard in the library domain. For both metadata standards LIDO and MARC, there are modules available in ECK that transform to EDM. Metadata managed and selected with ECK-ready systems, can be transformed to these formats with one mouse click.

According to good practice, data suppliers should indicate in their data if the content shared is under copyright and if yes, under which licence it can be made available (both strategically desired as well as legally allowed). Creative Commons[16] is the most popular type of licences used and also prescribed by Europeana. Adding licensing information

---

13 http://pro.europeana.eu/edm-documentation

14 http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/what-is-lido/

15 www.loc.gov/standards/marcxml/

16 www.creativecommons.org

to digital objects and the assets related[17] is a functionality included in any ECK implementation.

Step 4   VALIDATING FORMATS

Validation is a test to see if the transformation is correct and thus acceptable for the aggregator. Content providers would be interested to know if the transformation is successful. For this, a preview of the record is useful.

The ECK validation and preview can be used upfront from within the collection management system. The validation service of the ECK gives clear feedback and offers a preview of how metadata will be displayed in Europeana. This way, if needed, the user can go back one step, make corrections and prepare records that failed validation even before an export of data from the system or before the aggregator is involved.

Step 5   SUPPLY

Once data has been proven to be correct, it can then be handed over to the aggregator. Depending on the collection management system used, the ECK supports harvesting and/or data push.

Data push is more secure, more stable and most importantly it is the provider that takes initiative as opposed to harvesting where the aggregator collects the data. Data push can also support more direct feedback about the

---

17 A digital asset can be a picture or any file that represents the object digitally.

transfer of the data and its subsequent status (see module Set Management above).

This step includes incremental supply. This means that only records that have been changed are transferred: new data, updated fields, deletions. At the moment, Europeana only supports non-incremental harvesting. Using the ECK, the advantages of data push only work for data supply to an ECK aggregator. The only way to update a record set in Europeana is to re-supply all data.

If both the collection management system and aggregator implement ECK, a reversed supply channel can be introduced that allows a CMS to receive information from the aggregator. This functionality is called re-ingestion and has been tested in the framework of the Europeana Inside project fostering the enrichments on supplied data by Europeana[18]. In the future enrichments and user generated content is expected to provide added value to the collection owners and can provide an important business case to support the sharing of collections.

18 A Summary of the results of this pilot can be read here: http://pro.europeana.eu/pro-blog/-/blogs/cultural-heritage-institutions-re-ingesting-enriched-metadata

FUNCTIONAL DESCRIPTION OF THE ECK

To complete this ECK Guide we provide this technical chapter, for those technically versed and curious about how the ECK is built.

For this chapter we have selected and included the description of ECK components. A full technical documentation can be accessed at:

http://www.europeana-inside.eu/site/showfile.php?a=257

The ECK components are open source and free, all modules and full technical documentation can be accessed at:

www.europeana-inside.eu/eck

ECK Core

The ECK Core is a simple gateway which accepts data from Collection Management Systems (CMSs) and redirects it through the various ECK modules. Use of the Core module allows developers to direct output to a single point rather than ECK modules individually. ECK Core has been integrated into various aggregators.

CMS ECK supporting functionality

This functionality is provided within the CMS. It includes functions such as selecting and exporting records in the relevant metadata profile to an aggregator, reporting on the state of metadata through the process. In general, end users should be able to perform all the required tasks by interacting only with their CMS which incorporates this functionality.

Metadata profile definition

This module allows for the provision of multilingual alternative labels for use by the source CMS when displaying information to end users. It has been designed to be used in two circumstances:

–   To define labels for metadata elements. For example in the profiles of LIDO, MARC or EDM.

–   To enable the look-up of meaningful multilingual error definitions and guidance on steps to take.

PID Generation

This module creates persistent identifiers for those records which do not already have suitable identifiers. PIDs are a combination based on the institution URL (or some other identifier if no URL is available), record type and record accession number (or other suitable local identifier). This service allows links to be made between the identifiers used in the management system (CMS IDs) and other external identifiers (e.g. IDs generated by Europeana).

Mapping

Data mapping is the procedure by which a record is mapped from one format to another using the supplied rules. The mapping module has been developed for use where this functionality is not included in their CMS as part of the ECK supporting functionality.

Transformation

The transformation module transforms a record from one metadata schema to another. In the context of supply to Europeana, it can be used by a CMS or, more often, an Aggregator to transform data from a format such as LIDO, MARC or EAD to EDM, the metadata schema used by Europeana. Other transformations are possible. The transformation module can handle single XML files or several XML files in a ZIP file.

Preview

The preview generates a preview of a web page that simulates how the user's data will look when imported into Europeana (and optionally other targets e.g. intermediate aggregators, etc.).

Validation

During validation, certain fields of a record or a set, as well as the media files are checked for completeness and correctness:

– Checks metadata for conformance against the specified metadata schema and profile,

– Checks that media exists if referenced and is of a suitable size to fit with Europeana guidelines[19],

– Performs other checks on field contents for things that are not explicitly specified in the schema or profile (e.g. it assures that URI fields contain URIs and optionally, that titles are not too long).

Set Manager

The set manager is integrated in the CMS and allows a controlled way for records to be supplied to the aggregator. It also provides an OAI-PMH repository from which aggregators can harvest records as well as data push functionality.

The set manager allows for several actions (supply, validation, preview…) to be performed for a group of records at once and keeps the records up-to-date tracking changes performed in the CMS.

There are two types of sets, the working set in which changes can be made from the CMS and the live set which is in use by the aggregator. The working set becomes live set when the supply is performed.

---

19 Europeana Portal Image policy:
http://bit.ly/europeanaImageGuidelines

Aggregator ECK supporting functionality

This module provides the aggregator with a workflow for harvesting or receiving records from a CMS and returning to the CMS statistical information about the status of the records (number of records accepted, pending, rejected, published, etc.). It can, as well, validate records, transform to EDM and making them available to Europeana.

This functionality capacitates the aggregator to retrieve the published records in Europeana and to generate an enrichment record that can be requested by the CMS. These enhancements can include Europeana identifiers, concepts, places, subjects, agents or time expressions.

Statistics

This functionality related to the management of records, provides information on the performance of each module.

CREDITS

ECK has been developed in the framework of the Europeana Inside project by the following organisations:
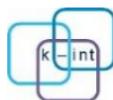
Find all information about the project and its partners under:
www.europeana-inside.eu