Grant Agreement 297292

EUROPEANA INSIDE

Summative Evaluation Report

Deliverable number	*D4.5 (v1)*
Dissemination level	*Public*
Delivery date	*May 2014*
Status	*Final*
Author(s)	*Nathalie Poot (KMKG)*

Revision History

Revision	Date	Author	Organisation	Description
v0.1	2014-05-23	Nathalie Poot	KMKG	Draft
v1.0	2014-05-28	Nathalie Poot	KMKG	Final version – review all partners

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Contents

1   Introduction

1.1 Background and role of the deliverable in the project

This report is part of Work Package 4 (WP4). This Work Package is dedicated to the **coordination of content** to Europeana: more than 960,000 records will be delivered using the Europeana Connection Kit (ECK).

This document is the first of two deliverables that present the outcomes of *Task 4.4 Evaluation of Content Strand*. Under this task ECK functionalities were implemented in the systems of the content partners (CPs) and an export of the content to Europeana was executed.

Both reports will give a summative **evaluation of the content delivery process to Europeana using the ECK**. *D4.5(v1) Summative Evaluation Report* (M26 – May 2014) presents the results of the initial export. *D4.5(v2) Summative Evaluation Report* (M29 – August 2014) is an update of *D4.5(v1)* and will present the results of the full content export from all content partners using the ECK.

The two versions of *D4.5 Summative Evaluation Report* (*v1* in M26 and *v2* in M29) are part of the **iterative development** plan. According to the Description of Work (DoW), WP4 was to start in M15 (June 2013) and end in M21 (December 2013). However it quickly became clear that the development schedule of the ECK as proposed in the DoW was unrealistic and changes were required to be able to follow a more agile approach as is commonly used in software development. A new development schedule has been drafted and takes into account four iterations of the ECK. For WP4, CPs test and report on each of these iterations after their release.

WP4 is further **dependent on the outputs of WP2, WP3 and WP5** for its deliverables. The development and production of the four releases of the ECK are part of WP3 (development) and WP5 (production). The deliverables for WP4 also build on previous reports within the work package.

The development of the ECK and the evaluation of iteration 3 are based on:
- *D2.1 Requirement Analysis*: explanation of all ECK requirements, based on a survey among the project partners.
- *D2.2 Use Cases:* three use case scenarios.
- *D2.3 Recommendations for Technical Standards:* research on best practice and quality instruments already in place within the Europeana project family.
- *D2.4 Functional Requirement:* there are three kind of requirements: high level requirements, workflow requirements and non-functional requirements. The workflow requirements are identified as: manage, select, prepare, validate, supply, data acceptance and enrich and return.
- *D2.5 Technical Specification:* describes the overall architecture of the ECK.
- *D3.5 Technical Integration Report:* progress report on the development of the ECK.
- *D4.2 Content Export Schedule:* presents the schedule for content delivery. It specifies the order in which participating institutions carry out the export of their data using the ECK.
- *D4.1(v1) Control Export Evaluation Report:* report on the test results from iteration 1 ECK prototype.
- *D4.1(v2) Control Export Evaluation Report:* report on the test results from iteration 2 ECK prototype.
- *D4.3 (v1) Export Evaluation Report:* report on the test results from iteration 3 ECK prototype.
- *D4.6 (v5) Technical Specification*: report on the technical specifications of the ECK.
- *D4.4 Content Re-ingestion Report:* report on the test results of content re-ingestion as part of iteration 3 ECK prototype.

The results of *D4.5(v1) Summative Evaluation Report* will be used for:

- *D4.5 (v2) Summative Evaluation Report:* an update of *D4.5(v1) Summative Evaluation Report*. This reports will provide a summative evaluation of the delivery process of all content using the ECK.
- *D4.3 (v2) Export Evaluation Report:* an update of *D4.3(v1) Export Evaluation Report.* It will report on the results of testing iteration 4 ECK, the production version.
- WP5: their object is to use the lessons learned in WP2, WP3 and WP4 to develop and launch a full production version of the ECK with accompanying support and documentation materials.

1.2 Approach

1. **Organising content delivery to Europeana**

   - **Practical guidelines for publication** on Europeana: how will CPs deliver their content to Europeana?

   - **Individual communication with CPs** to guide them through the content delivery process (email/phone/skype).

   - **Content monitoring**: close follow-up of the amount of records: 1) published on Europeana, 2) in the Inside Dark Aggregator (DA) and Culture Grid (CG), 3) harvested by Europeana and 4) published on the Europeana portal.

   - The content is monitored by the **content monitoring forms** that CPs complete each month and by **real time follow-up of the data in the DA and CG** (the total number of records, the number of invalid records, preview of the records).

   - **Cooperation with Europeana Ingestion Team**: meeting in The Hague in M22 (January 2014) and frequent email communication for content ingestion. Content ingestion is seen as an ongoing process.

   - **Content Partners Meeting**: at the 3rd Networking Event in Athens (M25 – 9th and 10th of April) the workflow to Europeana using the ECK was reviewed. The goal of the ECK was to simplify the process of contributing content to Europeana. How far are we in reaching that goal?

2. **Status content ingestion**

   - Workflow to Europeana: different steps were established for contributing content to Europeana.

   - Overview content delivery: how many records are published on the Europeana platform, how many are in the DA and CG and how many still need to be delivered?

1.3 Structure of the deliverable

The deliverable is structured in the following way:

- Organising content delivery to Europeana
- Status content ingestion
- Conclusions and next steps

Appendix I: Practical guidelines for publication on Europeana

2 Organising content delivery to Europeana

2.1 Practical guidelines for publication

Alongside testing and evaluating iteration 2 ECK and iteration 3 ECK CPs were asked to systematically supply content to the DA an CG[1]. To guide them through the content ingestion process, CPs were given practical guidelines (via Basecamp in M22 (January 2014) (Appendix I)[2].

2.2 Content monitoring

The focus of WP4 is to increase the quantity of high-quality material on the Europeana platform so that it can be made accessible to users. **More than 960,000 records** will be delivered within the Europeana Inside project using the ECK.

The full content ingestion in Europeana is taken place on a **staggered basis**. In *D4.2 Content Export Schedule* (M16) an ingestion plan was presented with the **overall structure** and **sequence** for **content delivery**. It specified the order in which participating institutions would carry out the export of their data using the ECK.

With only five more months to go in the project, content ingestion by each CP is monitored closely by WP4 lead. To make sure that content is published in time, there are three ways to follow up content ingestion:

1) **Monthly content monitoring forms:** completed and submitted each month by every CP in which they indicate:
   - The amount of records that was sent to the Dark Aggregator and Culture Grid
   - Other activities in regard to content delivery
   - Possible issues

2) **Individual support** on content delivery (email/phone/skype)
3) Real time follow-up of the content in the DA and CG

2.3 Content Partners Meeting

At the 3rd Networking Event in Athens a **Content Partners Meeting** was held in which the workflow to Europeana with the new ECK functionalities was reviewed. The goal of the ECK is to simplify the process of contributing content to Europeana.

How far are we in reaching our goal?

CPs responded to be overall satisfied with what has been developed and implemented in their system. After testing iterations 1 and 2 and with the start of testing iteration 3, they agree that it is possible to speak of a (partly) automated process to deliver data to Europeana[3].

---

[1] The results from testing iteration 1, iteration 2 and iteration 3 ECK are presented in *D4.1(v1) Control Export Evaluation Report* (M16), *D4.1(v2) Control Export Evaluation Report (M20)*, *D4.3 (v11) Export Evaluation Repor*t (M26) and *D4.4 Content Re-Ingestion Report* (M26).
[2] The content delivery steps correspond with the illustrative workflow described in *D4.6 Technical Specification* (K-INT): 10-11.

Most frequent mentioned STRENGHTS:
- Validation: the ECK makes it possible for CPs to validate their LIDO xml before the content is supplied to DA or CG.
- LIDO xml: CPs are now able to export their content in a valid LIDO xml. Not all CPs were able to do so before.
- Content delivery: for iteration 3 content partners will be able to directly push their data in the (dark) aggregator

Mentioned WEAKNESSES:
- Not all required functionalities are already tested and implemented in the system of the CP.
- Validation and preview modules only work on LIDO and not on EDM.
- Some CPs are struggling with LIDO xml. It is not always clear which field in their CMS corresponds with which field in LIDO.

---

[3] The results from testing iteration 1, iteration 2 and iteration 3 ECK are presented in *D4.1(v1) Control Export Evaluation Report* (M16), *D4.1(v2) Control Export Evaluation Report (M20)*, *D4.3 (v11) Export Evaluation Repor*t (M26) and *D4.4 Content Re-Ingestion Report* (M26).

3   Status content ingestion

3.1 Workflow to Europeana

All CPs are following the same workflow to deliver their content to Europeana using the ECK functionalities:

Step 1: Manage

ECK functionalities of data management and user management give CPs the opportunity to follow up on the exported content (e.g. which records have been exported, which records were altered when and by whom,…).

Step 2: Select

CPs make a selection in their system of the records they wish to publish. ECK functionalities provide flexibility in the selection of the records (selection of multiple records, selection based on specific values in a variety of fields, saved queries,….).

Step 3: Prepare

To supply content to the DA or CG the selected records need to be exported in LIDO xml or EDM from the system form the CPs. CPs prepare their content for export. The new ECK functionalities give flexibility in the choosing of the fields (e.g. mapping can be edited, mappings can be saved for repeated use, the system reports on problems when applying the mapping,…).

Step 4: Validate

Before CPs supply their data to the DA or CG, ECK functionalities make it possible to validate their LIDO xml in their own system. If the LIDO xml is not well formed their system will report on irregularities (e.g. if mandatory LIDO fields are missing). Based on the feedback, CPs correct the invalid LIDO records in their system and re-export.

Step 5: Supply

CPs supply their data to the DA or CG by either OAI-PMH or push (SWORD). If there is an error in the data, the system is able to re-start the supply process, only for the failed records.
All data is validated again in the DA and CG. Records that are successfully validated are transformed to EDM.

At that moment the content is also checked by the WP4 lead who informs the CPs if corrections are needed (quality of the metadata e.g. if there are odd signs in the fields, like # , if content is repeated in multiple fields,….).

The DA and CG support incremental harvesting. If corrections are needed in data that is already in the DA or CG, then CPs only need to supply the adjusted data again.

Step 6: Data acceptance

Once all records are valid in the DA and CG, harvest instructions are send to Europeana. Based on the feedback given by Europeana, WP4 lead reports back to the CPs if adjustments are needed on their data.

It is however not possible that the system of CP can withdraw earlier delivered records instantly from Europeana or that the system can keep the data that are already in Europeana-up-to-date. If an update is needed, then the content need to be re-ingested by the Europeana Ingestion Team.

3.2 Overview content delivery

From December 2013 until May 2014 a rich variety of objects have been published on the Europeana portal by different CPs:
- Greek and Foreign fine and decorative art and sculpture (NAG)
- Photographic Archives, Neohellenic Architecture Archives, The Nikos Hatzikyriakos-Ghika Gallery collection of paintings (BEN)
- Egyptian Art, Classical Antiquities, Old Master Paintings, Sculpture, Prints and Drawings, Department of Art after 1800 (FAB)
- Photographic collections, Iconographic collections and Objects (SEI)
- Cultural History Objects; musical notes (handwritten); scientific reports and papers (archeology and etnography); printed articles (cultural history) (SLV)
- Collection of Social History Images from several UK cultural organisations (HIM)
- Collections of the Imperial War Museum (CT)
- Decorative arts, lace, costume and accessories, tapestry, Japanese prints (KMKG)

In the following months data from different CPs will be added:
- Christian history and heritage in Belgium (KADOC – KU Leuven)
- Highlights from different SPK collections (SPK)
- Photos, Postcards fauna, posters from exhibitions IRScNB, ethnographic photos… (RBNIS)
- Selection from the Archaeological Department, the Historical Repository, Historical Photo Collection, Hungarian Historical Gallery and the Coins Collection (HNM)
- Art and Relics Collection, Library collection and Manuscript collection (PIM)

<u>Planning content ingestion</u>

In M22 (January 2014) a meeting was held in The Hague with Europeana in which the content ingestion process was discussed. Content ingestion was seen as an ongoing process: as soon as content is ready in the DA or CG harvest instructions would be send to Europeana to harvest the data.

There are 2 milestones for content ingestion for WP4 (DoW)
- Year 2: 100,000 records published on Europeana (December 2013)
- Year 3: more than 960,000 records published on Europeana (end of the project in September 2014).

Content provider	DoW	Aggregator (DA or CG)	Published on Europeana
KADOC KU Leuven (BE)	10.000	0	0
Collections Trust - CT (UK)	590000	34983	34983
Stiftelsen Lansmuseet Vasternorrland - SLV (SE)	130030	48200	57349
Municipio do Seixal - SEI (PT)	17000	12149	12149
Xantys Limited / House of Images - HIM (UK)	100000	2941	2941

Magyar Nemzeti Múzeum - MNM/HNM (HU)	30000	30620	0
Szepmuveszeti Muzeum - FAB (HU)	8000	4076	4076
Benaki Museum - BEN (GR)	13000	13018	13018
Stiftung Preussischer Kulturbesitz - SPK (DE)	10000	10760	0
Institut Royal des Sciences Naturelles de Belgique - RBINS (BE)	3000	0	0
Royal Museums of Art and History - KMKG (BE)	30000	18313	14826
National Gallery-Alexandros Soutzos Museum - NAG (GR)	9000	5863	3032
Petofi Irodalmi Muzeum - PIM (HU)	10000	10000	0
Dr. Kovacs Pal County Library (HU) - associate partner	0	41470	0
TOTAL	**960.030**	**232393**	**142374**

Comments:

- KADOC KU Leuven (BE) and Institut Royal des Sciences Naturelles de Belgique - RBINS (BE): their content is ready. They will supply their data to the DA directly in EDM.
- Municipio do Seixal - SEI (PT): will make the remaining 4,851 records available to the DA the 30th of May.
- Stiftelsen Lansmuseet Vasternorrland - SLV (SE): have supplied 48,200 records to the DA. Harvest instructions will be send to Europeana.
- Magyar Nemzeti Múzeum - MNM/HNM (HU) and Dr. Kovacs Pal County Library - kkmk (HU) - associate partner: all content is supplied to the DA. Harvest instructions are sent to Europeana.
- Szepmuveszeti Muzeum - FAB (HU): will make the remaining 4,000 records available to the DA the 1ste of June.
- Stiftung Preussischer Kulturbesitz - SPK (DE): all content is supplied to the DA. Harvest instructions are sent to Europeana.
- Royal Museums of Art and History - KMKG (BE): will make the remaining 11,678 records available to the DA the 30th of May.
- National Gallery-Alexandros Soutzos Museum - NAG (GR): will make the remaining 3,137 records available to the DA the 30th of May.
- Petofi Irodalmi Muzeum - PIM (HU): all content is supplied to the DA. Harvest instructions are sent to Europeana.
- Xantys Limited / House of Images - HIM (UK): will be adding each of their other 8 datasets (totalling approx 100,000 records) in May.
- Collections Trust - CT (UK): Imperial War Museum is providing approx 2000 records each month.

3.3 Remaining issues

While overall CPs are satisfied with the current workflow, some minor adjustments related to the LIDO-EDM mapping would improve the content ingestion process.

Validation and preview

At the moment not all data that is successfully validated by the validation module in the system of the CP or in the DA or CG, is also validated by Europeana. The system should ensure that successful validation means validation by Europeana at ingestion as well. There are for example mandatory fields missing e.g. missing of edm:rights, or dc:subject or dc:type or dc:coverage or dcterms:spatial.

At the moment both the validation and preview service only work on LIDO and not on EDM. Consequently CPs that supplying content available to the DA or CG in EDM, cannot be validated or previewed.

Identifiers

Currently the identifiers in LIDO are all mapped to dc:identifier and published on Europeana. For CPs there is no need for all the identifiers to be displayed.

The structure of PIDs was defined within EUinside as: "creation of PIDs based on institution URL (or some other identifier if no URL is available), record type and record accession number (or other suitable local identifier), combining these into a PID". However, once published on the Europeana portal, Europeana forms it into a clickable link that leads to nowhere.

Conclusions and next steps

The goal of the ECK is to simplify and automate the process of contributing content to Europeana. The ECK services developed and implemented by the technical team in the systems of the CPs belong to all the workflow steps to Europeana: manage, select, prepare, validate, supply, data acceptance and enrich and return.

For delivering their content, all CPs are using ECK functionalities. At the CPs meeting in Athens, CPs stated to be overall satisfied with the new functionalities that they tested as part of iterations 1 and 2. With the start of testing iteration 3, CPs agreed that it is possible to speak of a (partly) automated process to deliver data to Europeana.

What has been achieved:

- All CPs are satisfied with the new ECK functionalities for delivering content to Europeana.
- 142,374 records are published on Europeana.
- The remaining data is already supplied to the DA and CG and is ready to be harvested by Europeana or the data will be ready in the following month.

Next steps:
- Continuation of monitoring and guiding content delivery in the DA and CG.
- Continuing to send harvest instructions to Europeana for content ingestion.
- Providing feedback to the CPs on their data.
- To reach to goal for WP4 by the end of the project: to publish more than 960,000 records on the Europeana portal using ECK functionalities.

APPENDIX I: Practical guidelines for publication on Europeana

For successful content delivery, CPs were asked to:

1) Export the records from your CMS into a <u>valid LIDO xml</u>

2) Check your LIDO xml <u>before</u> you make your data available to the (dark) aggregator. If records are invalid, correct them with help from your technical partner. This will avoid a re-upload in the (dark) aggregator.

3) Make your data available to the (dark) aggregator: you have received a <u>login</u> and a <u>username</u> from Phill Purdy to upload (web form), harvest (OAI-PMH) or push (SWORD) your records in the aggregator (DA or CG). If this is not the case (for example for associate testing partners), please contact Phill Purdy (phill@collectionstrust.org.uk).

   (<u>please note that</u>: upload by web form is only available as a temporary work around to enable WP4 targets to be met. By iteration 3 all technical partners should have implemented either data push (via SWORD) or OAI-PMH to meet the functional requirements of the ECK).

4) Make your data available to the (dark) aggregator.

5) Contact Phill Purdy (phill@collectionstrust.org.uk) and me (n.poot@kmkg-mrah.be) when your records are successfully in the (dark) aggregator so we can monitor them.

6) Once your data is processed, the following status of your records will be given: it will be marked how many records:
   - Failed validation (are invalid)
   - Have been processed successfully and were transformed to EDM (are valid)

7) <u>Invalid records</u> need to be adjusted. You will need to go back to your CMS, adjust the mapping and re-upload the records in the (dark) aggregator.

8) Once your records are successfully transformed into EDM, they will be harvested by Europeana.

For the <u>amount of records</u> you need to deliver and the <u>deadline:</u> please respect the <u>ingestion plan</u> (see deliverable D4.2 Content Export Schedule, appendix II).

➢ I will contact each content partner individually to follow-up on the amount of content that needs to be in the aggregator at the indicated time.