3D Digitisation of Icons of European Architectural and Archaeological Heritage

3D ICONS

D6.2: Report on harvesting and supply

Authors:
Dimitris Gavrilis (Athena R.C)
Eleni Afiontzi (Athena R.C)
Dimitra-Nefeli Makri (Athena R.C)
Athanasios Tsaouselis (Athena R.C)
Nikolaos Kazakis (Athena R.C)
Christodoulos Chamzas (Athena R.C)
Sheena Bassett (CISA)

3D ICONS is funded by the European Commission's
ICT Policy Support Programme

cip competitiveness and innovation framework programme

Revision History

Rev.	Date	Author	Org.	Description
0.1	07/01/15	Dimitris Gavrilis, Dimitra-Nefeli Makri	Athena R.C. (DCU)	First draft for review
0.2	09/01/15	Eleni Afiontzi	Athena R.C. (DCU)	Corrections
0.3	12/01/15	Dimitris Gavrilis	Athena R.C. (DCU)	Corrections
1.1	26/01/15	Nikolaos Kazakis, Athanasios Tsaouselis C. Chamzas	Athena R.C. (CETI)	Metadata Quality Assurance Checks, Progress Monitoring Tool 3D-Icons Portal Harvesting
1.2	28/01/15	S. Bassett	CISA	Corrections
2.0	29/01/15	C. Chamzas, S. Bassett	Athena R.C. (CETI) CISA	Corrections
3.1	13/02/15	C. Chamzas, S. Bassett	Athena R.C. (CETI) CISA	Additional of partners figures, final corrections.

Revision: 3.1[Final]

Authors:

Dimitris Gavrilis (Athena R.C)
Eleni Afiontzi (Athena R.C)
Dimitra-Nefeli Makri (Athena R.C)
Athanasios Tsaouselis (Athena R.C)
Nikolaos Kazakis (Athena R.C)
Christodoulos Chamzas (Athena R.C)
Sheena Bassett (CISA)

Contributors:

Andrea D' Andrea (CISA)

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

3D-ICONS is a project funded under the European Commission's ICT Policy Support Programme, project no. 297194.

Contents

Executive Summary

The goal of the 3D-ICONS project is to create and supply 3D models and related content regarding architectural and archaeological masterpieces of European and world cultural significance to Europeana. This required the development of a metadata schema suitable for recording all the information associated with 3D models. The existing CARARE schema (which is specifically designed for archaeological content) was updated and mapped to EDM to enable the content providers to create metadata that could be ingested by Europeana, thus providing public access to their models and associated videos and images. The same ingestion infrastructure was used as developed in the CARARE project, updated to accommodate the changes in the schema. However, most of the 3D-ICONS partners were not experienced with creating metadata so a new tool, the Metadata Editor was created to assist the content providers to create metadata from new quickly and easily without technical knowledge of XML or the CARARE2 schema. Four of the partners who held existing metadata used MINT2 to map and ingest their data into MORE2.

This deliverable describes the harvesting system for the 3D-ICONS project, which allows content providers to provide their content to the system using a methodology which will allow for regular automated harvests. The specific technologies employed for the communication between the separate modules, the mapping of CARARE2 to EDM and the ingestion into Europeana are discussed in detail along with how the process was managed and monitored. The quality control, a highly important part of the ingestion process, is also explained.

The problems encountered, tools developed and solutions adopted during the whole procedure are described. These also include the 3D-ICONS Portal which proved an invaluable tool for the checking metadata ingested into MORE2 as the complete set of CARARE2 data fields could be inspected (not just those fields that get mapped to EDM). Last but not least, statistics about the items delivered to Europeana as well the type of each of them are presented. Progress since May, when the first records were sent to Europeana, is shown graphically with the ingestion status at 31st January 2014. The report concludes with a summary of the main issues experienced by the content providers, planning for the following months as further ingestion continues (mainly corrections plus a few additional 3D models that partners are able to contribute) and the overall performance of the Project with respect to the numbers of items defined in the DoW.

1. Introduction

This document describes the process of harvesting and the provision of metadata to Europeana by the 3D-ICONS project. There are three main components that can used for the harvesting. These are the Metadata Editor, MINT and MORE2 aggregator. Deliverable 4.2 “Interim Report on Metadata Creation” presents in MORE detail the functionality of these three systems.

This report begins with a general overview of how the content is harvested (Section 2) and the technologies used. Section 3 describes the ingestion process and how the project organized the delivery of their metadata for ingestion. Section 4 discusses the individual tools used by the project and their role in the ingestion process. Section 5 presents the Metadata Quality Control Process and how quality control was implemented throughout the ingestion process and includes the statistics of the content sent and published to Europeana. Last but not least, problems encountered and the solutions applied are discussed in Section 6 along with figures and graphs which illustrate the project’s metadata ingestion progress over the last few months. This section concludes with the planning for ingestion in February and March 2015 (after the end of the project). Section 7 is the conclusions. Further information is provided on the ingestion submission packages, the Metadata Editor and MORE2 in the Annexes.

2. Harvesting the content

The first step for harvesting is the user to create their metadata, following the guidelines from the Deliverable 4.2. The content provider can create metadata through either the Metadata Editor or MINT or by using both systems. Most of the content providers in 3D-ICONS had no metadata for their digital content and so a new tool, the Metadata Editor, was created to simplify input through a series of forms with inbuilt error checking which ensured that records had to contain the requisite information to be mapped to EDM. For the remaining partners, three used MINT 2 to perform direct mappings between their own repositories (having previous experience of ingesting metadata in to Europeana) and one other used a combination of methods to export information from a museum repository, map this MINT 2.0 and complete the required fields in the Metadata Editor.

After completing this procedure, the package is ready for ingestion to MORE (Figure 1).

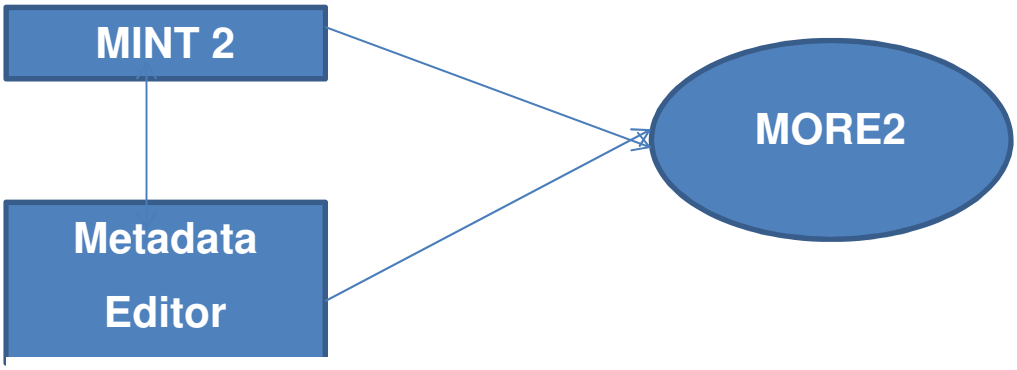


Figure 1: Harvesting the content

Repository communication with schema mapper

After the content providers have finished mapping their metadata to the CARARE 2.0 schema either though MINT 2 or though the Metadata Editor, metadata for all digital objects is mapped and ingested into the repository (MORE2).

2.1 Method of communication

The ingestion mechanism relies on a set of REST-based web services. The web service allows the mapping tool, MINT 2, to ingest information into the MORE2 repository and only requires one parameter which is the URL of the submission package. The web service downloads the submission package, verifies it and ingests it into the repository.

The REST-based web service is available at http://3dicons.dcu.gr/ and has the following specifications:

- implement the GET method of HTTP request,

- accept the one variable (GET type) with name package. This variable will contain a proper URL of the location of the package to ingest (e.g.:http://3dicons.dcu.gr//ingest/index.php?package=http://194.177.192.14/carare/package1.zip)

After downloading and processing the submission package, the ingestion service returns an XML formatted response, a description of which can be found in Annex I.

2.2 MORE2 Repository export to Europeana

The MORE2 repository allows for its content to be mapped to EDM and exposed to Europeana. A XSLT schema is used for the metadata mapping mechanism to create an EDM datastream, which includes all the necessary information. Each transformed EDM schema is stored in the **EDM** datastream inside the repository. This way, the processing load on the server is reduced since the server only needs to create the EDM datastreams when:

- A new CARARE 2 datastream is ingested (either from the interface with the schema mapper (MINT) or directly from the repository).
- The existing CARARE 2 schema is edited directly in the repository.
- The CARARE 2 ➔ EDM mapping description is modified.

Note that (1) is the normal operating situation for MORE2. Updating the CARARE schema (2.) has not been required during the project as careful consideration was given to the addition of the metadata fields specific to 3D models and some additional refinements were also made at the time of implementation of CARARE 2 based on the experience of using the original schema in the CARARE project. However, some mapping adjustments (3.) were initially required as the CARARE 2 schema is complex and EDM has several options for the metadata required by Europeana.

An OAI-PMH provider exposes the contents of the repository to Europeana making possible for Europeana to harvest the provider's collections on demand.

3. Supply to Europeana

The ingestion cycle of Europeana is shown in Figure 2.

3D-Icons Diagram Timeline for Ingestion

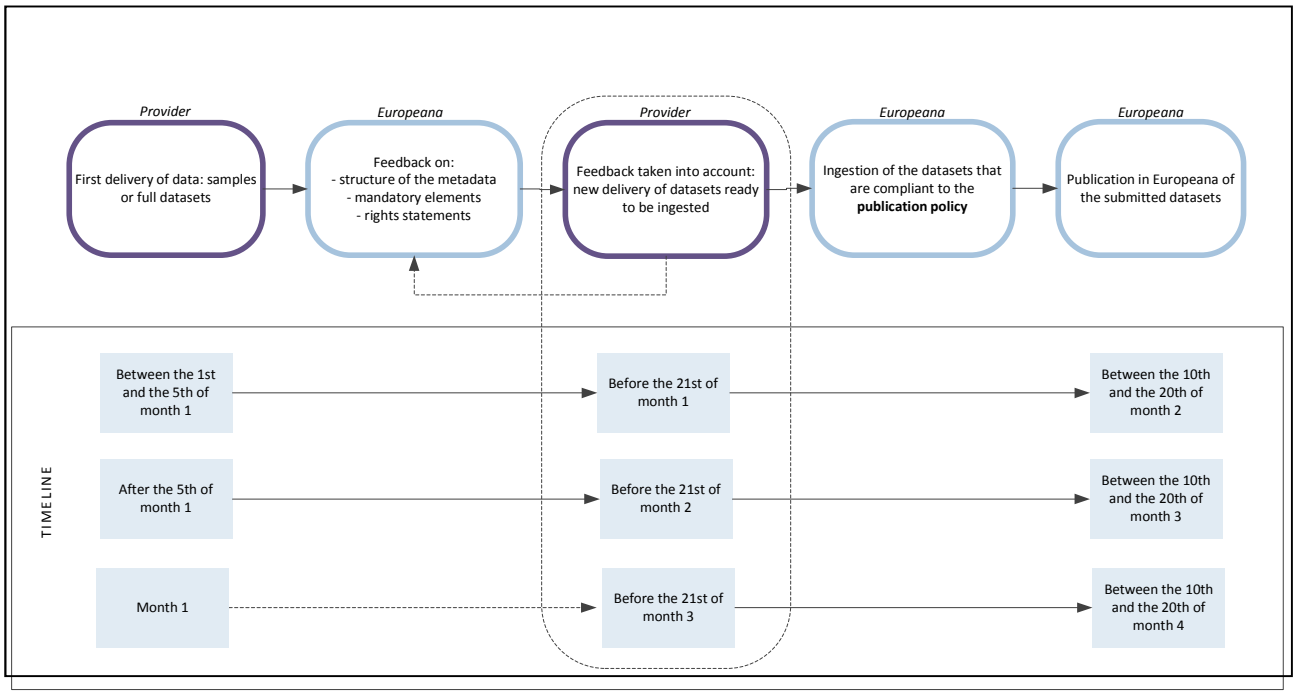


Figure 2: 3D-Icons timing diagram for ingestion

Metadata creation did not really start until Year 3 for most of the content providers. Following a workshop on the use of MINT 2 and MORE2, the need for a simplified metadata creation tool became apparent which led to the creation of the Metadata Editor during Year 2 and which following testing and trials by the partners, finally became ready for use at the start of Year 3. Following the project meeting in Jaen, March 2014, a metadata ingestion schedule was created for each partner with monthly targets set for the three types of content supplied, 3D models, images and videos (see D4.2 Interim report on Metadata Creation for further information). In accordance with the Europeana Harvesting Schedule shown in Figure 2, each partner was required to prepare a small sample set of metadata records by 15th May in order for Europeana to provide their initial feedback. Fourteen packages were submitted and the first feedback was provided by the Europeana Ingestion Team on the 18th July. As might be expected, several issues with the 3D-ICONS metadata were identified from the test batch and CETI liaised closely with each partner to explain how to resolve the problems. Some of these stemmed from an incomplete understanding of what was ultimately required by Europeana such as complete and compliant Rights Statements and the difference between the Rights regarding the metadata and the Rights for the digital content. The majority of the issues could be attributed to missing data, data in the wrong fields, duplicated identifiers and the like. The first batch of 423 EDM records (315 3D models) were published in the Europeana portal on the 5th August, nearly three months after the first ingestion deadline. After these, the

publication numbers grew steadily as the content providers gained experience and created better quality metadata. Guidelines were also issued to assist partners with this task.

4. Tools, Improvements and Issues Addressed

In order to facilitate the harvesting process, two new tools were developed for the 3D-Icons project. These are:

- Metadata Editor Tool
- 3D-Icons portal

In addition, a Metadata Quality Process was implemented to ensure that as many issues with the metadata were identified and corrected prior to ingestion by Europeana plus a post-publication independent check on the published records in Europeana was also carried out.

The MINT mapping tool and the MORE repository which were originally developed in the CARARE project, were upgraded to work with the CARARE 2 schema for the 3D-ICONS Project.

4.1 Metadata Editor tool

The Metadata Editor (http://3dicons.dcu.gr/metadataeditor/) has been designed and implemented to support and facilitate partners in the metadata creation. As it is already discussed (Deliverable 4.2), the tool is basically set up on the declarations of blocks or groups of information that are repeated for each record and that can be duplicated for new records to speed up the metadata creation process (as much of the data such as the organisation, contacts, technical data etc. is the same for all the records of an organisation). The tool provides a means for the end user to input their metadata without an in-depth technical knowledge of the CARARE 2 schema or XML tools, i.e. focussing on the information aspect without the need for technical expertise. In addition, extensive Help information is provided for each input field to guide the end user and the tool indicates the level of completeness of each record created, not permitting the publication of a record for ingestion into MORE2 until all the mandatory fields have been completed (see Annex II). The Metadata Editor does not perform checks on the "correctness" of the data entered – this would have been very complicated and time consuming to implement given the tight schedule for ingestion to be achieved by the project. Consequently, a manual checking quality control system was used instead (c.f. Metadata Quality Process).

4.2 The 3D-Icons Portal

The 3D-Icons Portal (http://3d_icons.ipet.gr/) was originally developed for the presentation of the 3D-Icons items on (i) a geolocation system and (ii) presenting the rich metadata provided by the CARARE 2 schema. However, it was soon used by the partners to check the validity of their data before it was published in EUROPEANA. The capability to publish data in the portal in a very short time after it was published in MORE2 was very essential for this operation. The partners were able to see their data published and decide if corrections were necessary before the final submission to Europeana. At the same time, if their data was already published in Europeana's portal, there is a direct link to it in the Portal so they can view it as it is presented in Europeana. Some relevant screens of the 3D _Icons portal are given in Annex III.

One further facility provided by the Portal was the statistical summary of the digital resources (by type) for each partner. This was useful for the partners for checking against the number of records published in MORE2 as it would identify any issues if there was a mismatch. It allowed the content providers and the project management to track the monthly ingestion targets specified in the Ingestion Schedule and to report the ingestion progress to the Commission. Although the number of metadata records created and published was intended to be tracked in the Progress Monitoring Tool, this function was superseded by the Statistics report in the Portal as the numbers (and type of digital resource) are extracted automatically from the metadata published to MORE2, providing a greater level of detail and saving time as no manual input of the figures is required.

4.3 Metadata Harvesting from MORE2

The process of metadata harvesting explains the way that the 3D-Icons Portal manages to harvest metadata for web representation. The 3D-Icons Portal communicates via an API with MORE2 to fetch metadata for each organisation providing content in the project. A specific parser was implemented to manipulate the response of the API. The parser gets the xml response from the API and, after the collection of all the required data, imports this data to the 3D-Icons Portal for the web representation.

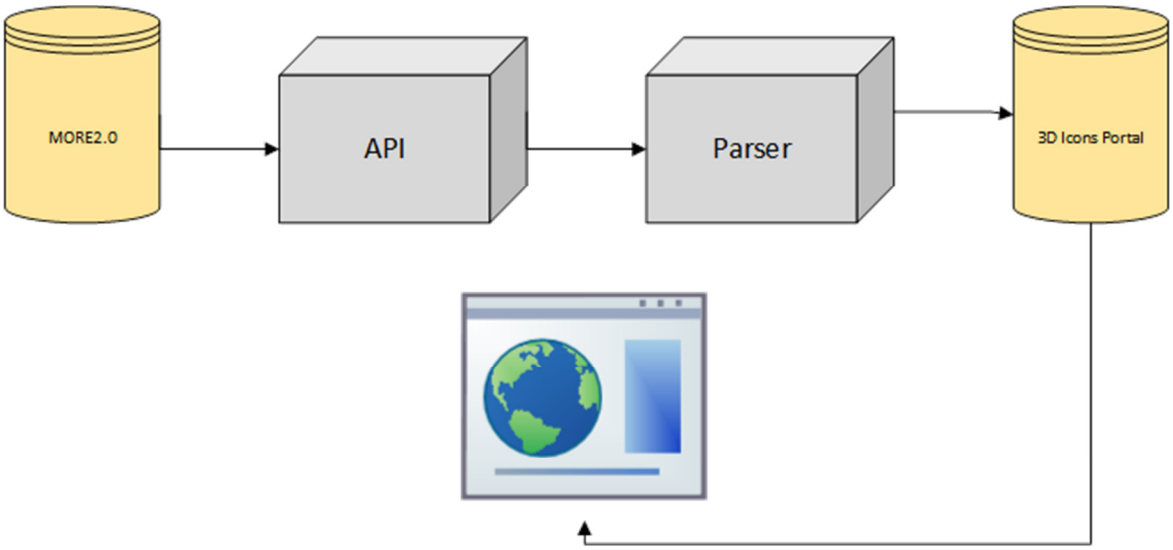


Figure 3 . MORE2.0 API – 3D-Icons Portal relation

The whole process is described in Figure 3. Organizations publish their metadata packages to MORE2 that were ingested through either Metadata Editor or MINT 2. Provided that packages have been successfully ingested to MORE2 and characterized as published, the metadata of the packages are ready to be parsed by the 3D-Icons Portal Parser. The parser is executed in given time intervals. Before every official harvesting from Europeana, the parser runs repeatedly and ensures the efficient validation of metadata. This pipeline offers the advantage that the partners are able to validate both the semantic content of the metadata and the syntax errors. After every execution of the parser into the 3D-Icons Portal, full statistics of the metadata packages are generated per organization. (http://3d_icons.ipet.gr/index.php/statistics) (Figure *9*).

5. Metadata Quality Control Process

Metadata was a new concept for most of the partners in the 3D-Icons project. Two training workshops, one in Xanthi (June 2013) and one in Marseille (October 2013), were organized for training on the CARARE 2 schema, metadata ingestion and harvesting. Even if grammatical and syntactical errors were identified by MINT 2 and MORE2 during the ingestion process, and then by Europeana during the harvesting process, it was soon realized that a Metadata Quality Assurance Check (MQAC) should be implemented to check the quality of the supplied digital resources and metadata. Two control points, MQAC and EQAC, were introduced; one during the harvesting process of metadata by MORE2 and a second one after the data were published into Europeana.

5.1 During the harvesting process (MQAC)

The scope of the Metadata Quality Assurance Checks (MQAC) is to verify the quality and integrity of the metadata (e.g. geographical coordinates, 3D-model link etc.) of all items of the packages ingested into MORE2 *before being harvested by EUROPEANA*. Although some checks of the metadata are made by MINT 2 and MORE2, these mainly focus on the identification of Grammatical and/or Syntactical errors of the inspected metadata. As a result, a more thorough investigation of the metadata reliability provided by users is imperative due to the high level of lack of experience with metadata. The area of interest for the MQAC is indicated in Figure 4, which depicts the flow chart of the *Progress Monitoring Tool*.

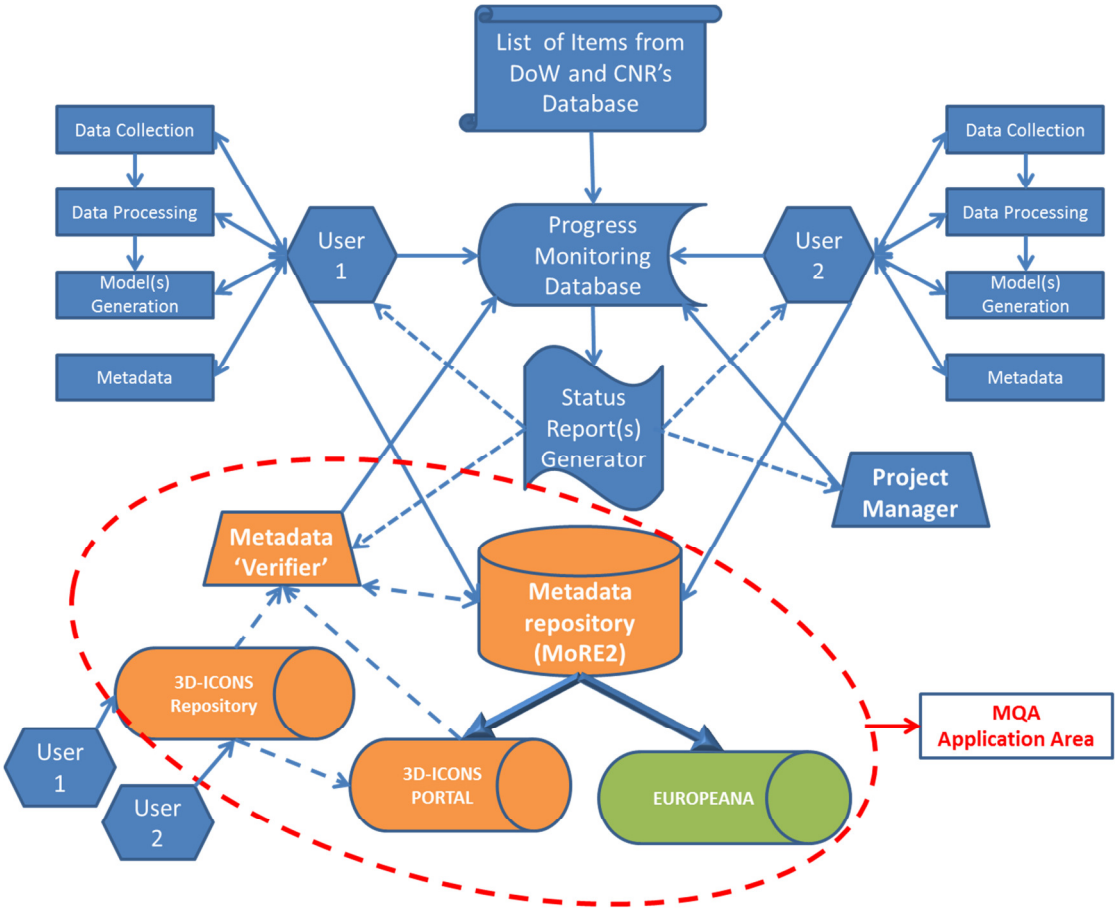


Figure 4. Flow chart of Progress Monitoring Tool; the red circle depicts the area of interest where Metadata Quality Assurance (MQA) is applied.

For this purpose, a ***manual inspection*** of the most significant metadata fields whose accuracy is the key element for the efficient appearance of each Item and the fulfillment of the scope of the 3D-ICONS project takes place. That is:

- the ***geographical coordinates*** of the Item, which provide the most accurate information about its location,
- the ***URI of the 3D-model,***
- the ***URI landing-page,*** and
- the ***thumbnail*** of the 3D-model provided by each User,

These are called MQA Items hereafter.

In brief, MQAC's **scope** is **only** to check if the MQA Items are valid and correspond to the model described for all available 3D-models and **not** to trace any errors regarding the mapping of the metadata and/or the 3D-model quality.

More specifically, as soon as a Package is classified as publishable by a data provider and waits to be harvested from EUROPEANA, it is also harvested from the 3D_icons portal (http://3dicons.ceti.gr). As a result the "published" metadata in MORE2 are available and are checked for errors (*Figure 5*).

Heritage Asset ID	Organisation	Title	QuickPoints	DR Title ID	Link	Thumbnail	Landing Page
http://3dicons.dcu.gr/object/HA/303	Athena R.I.C., ILSP, Xanthi, (former Cultural and Educational Technology Institute (CETI))	Ιερός Ναός Παναγίας Αχειροποίητου (Φωτογραφίες)	40.63490 22.94789	Image #1 of the Church of Panagia Acheiropiitos http://3dicons.dcu.gr/object/DR/332	http://3d_icons.ipet.gr/3dicons/images/ARC3DICONS_3D_1ACH/ARC3DICONS_3D_1ACH_1_1.jpg	http://3d_icons.ipet.gr/3dicons/thumbnails/ARC3DICONS_3D_1ACHT1.jpg	http://3d_icons.ipet.gr/3dicons/index.php?id=ARC3DICONS_3D_1ACH&res=L#imagesTab
http://3dicons.dcu.gr/object/HA/121	The Discovery Programme	3D Model of St Saviours Priory, Glendalough	53.007913 -6.312195	Perspective view4 of untextured 3D model of St Saviours Priory, Glendalough http://3dicons.dcu.gr/object/DR/1849	http://3dicons.ie/images/CONTENT/SITES/GLENDALOUGH/SAVIOURS/GLEN_SAVIOURS_ISO4.jpg	http://3dicons.ie/images/mgthumbnails/268x200-images-CONTENT-SITES-GLENDALOUGH-SAVIOURS-GLEN_SAVIOURS_ISO4.jpg	http://3dicons.ie/index.php/3d-content/sites/38-st-saviours-glendalough#images

Figure 5. Snapshot of the metadata checked during the MQA using the 3D-ICONS-Monitoring Progress Tool (http://3dicons.ceti.gr/audit/)

At this point, a manual control takes place through which it is confirmed whether all required metadata fields are completed for all Cultural Monuments and their Cultural Entities. At the same time, the URIs of each Item, namely for the landing-page (digital resource) of the 3D-model and/or the 3D-model itself and for the image to be used as thumbnail are checked for broken links, that they indeed lead to the correct model for each Entity in accordance with the description of the content provider. In addition, the geographical coordinates of the Entity are also evaluated for correspondence to the correct location of the Item.

In the case of incomplete metadata and/or invalid URIs of one or more 3D-models, the provider is asked to make corrections accordingly. If inaccurate metadata is detected, then again the provider is asked to make appropriate amendments. In this case, a mail is sent to the data provider which informs them about the problematic Entities and the issues to be solved (*Figure 6*) along with a reference guide which includes general instructions on how to tackle the problem (Figure *7*).

a/a	Heritage Asset ID	Title	DR Title ID	Landing Page	Notes
1	http://3dicons.dcu.gr/object/HA/835	Cavalieri Square - photographic documentation	Cavalieri Square - Photo - cavalieri_im_a101	http://vcg.isti.cnr.it/europeana/cavalieri_square/landing/cavalieri_square_cavalieri_im_a101.html	wrong landing page
			ISTI_CNR_CAVALIERI_cavalieri_im_a101		
2	http://3dicons.dcu.gr/object/HA/835	Cavalieri Square - photographic documentation	Cavalieri Square - Photo - cavalieri_im_a102	http://vcg.isti.cnr.it/europeana/cavalieri_square/landing/cavalieri_square_cavalieri_im_a102.html	wrong landing page
			ISTI_CNR_CAVALIERI_cavalieri_im_a102		

Figure 6. Sample of the information sent to the data provider after the MQA indicating the insufficient data

Issue	Explanation	Action	Path for the corresponding CARARE2 Schema field
wrong QuickPoints (coordinates)	Object's coordinates lead either to different place than the appropriate one (do not comply with the object's source) or nowhere (e.g. due to erroneous use of decimal separator-point or comma)	Check and correct the longitude and latitude values appropriately in the corresponding field	HeritageAsset → spatial → geometry → quickpoint → x → y
missing QuickPoints (coordinates)	Object's coordinates are completely missing	Fill with the described object's longitude and latitude values the corresponding field	HeritageAsset → spatial → geometry → quickpoint → x → y
wrong 3D-model Link	The 3D-model link is either dead or invalid	Check and correct the 3D-model link appropriately in the corresponding field	DigitalResource → link
missing 3D-model Link	The 3D-model link is completely missing	Fill with the described object's 3D-model link the corresponding field	DigitalResource → link
3D-model is not displayed properly	The 3D-model link is valid but the 3D-model itself does not appear	Check and correct the 3D-model associated files in the link	DigitalResource → link
wrong Thumbnail link	The Thumbnail link is either dead or invalid	Check and correct the Thumbnail link appropriately in the corresponding field	DigitalResource → object
missing Thumbnail link	The image illustrating the object's thumbnail is completely missing	Fill with the Thumbnail link the corresponding field	DigitalResource → object
wrong Landing Page	The Landing Page link is either dead or invalid	Check and correct the Landing Page link appropriately in the corresponding field	DigitalResource → isShownAt
missing Landing Page	The Landing Page link is completely missing	Fill with the Landing Page link the corresponding field	DigitalResource → isShownAt

Figure 7. List with the most probable results of the Metadata Quality Assurance Checks and proposed remedial actions by the Users

When an issue found during the MQA is not included in the most commonly found problems, the data provider is explicitly informed about its nature (e.g. using commas instead of points in the numbers of the geographical coordinates). Once the provider is informed about the deficient metadata, corrections are made and the metadata updated to be again classified as publishable.

In the case of correct and approved metadata, confirmation is given and the item returns to MORE2 to be harvested by EUROPEANA.

5.2 Quality Control on Europeana portal

Following the initial publication of the first batch of metadata records in Europeana, a review of the content revealed several issues with the metadata and the quality of some of the 3D models. Consequently, an independent manual quality check on 548 published records was carried out by a researcher in early November, who checked for the following:

- Title – did this accurately describe the object represented by the digital resource enabling it to be found when searched for? (Some partners had mapped museum reference numbers).
- Thumbnail – was this present and of good quality
- Landing Page – if specified, did the URL work? (A specific issue was identified where all URLs must be preceded by "http:// "in Europeana in order for them to work.)
- 3D PDF- was downloading enforced (most browsers are unable to display 3D PDF)?
- 3D QUALITY – did the image load well, was of good quality and the default view as expected?
- Description – was this adequate to describe the object? (Some metadata was very brief).
- Other metadata completeness – were the displayed fields complete and containing coherent data?

This exercise identified several problems which partners were able to able to rectify and resubmit their data.

The project manager also performed several checks following this review, working with individual partners to correct their metadata and also improve their landing pages so that these were more user-friendly to Europeana end users. It was especially important to test the metadata and 3D models using lower specification PCs with different browsers on standard household internet connections in order to replicate the end user experience. Most of the 3D-ICONS partners have very high specification computers and fast broadband connections and so were not in a position to perform these tests. Both CISA and CETI undertook this type of testing.

5.3 Ingestion and Harvesting Statistics

The content is delivered to Europeana in EDM through the OAI-PMH protocol. For each provider, a unique OAI-PMH URL is being sent making possible for Europeana to harvest the provider's collections on demand.

The following Table 1 shows statistics about the content that has already been delivered and published to Europeana as of 31/12/2014.

Data provider name & acronym	**Number of PCHO's**	**Number of webResources**
CETI - Athena Research and Innovation Center	44	357
CNR-ISTI- Instituto di Scienza e Tecnologie dell Informazione	72	1031
ICA- Interdepartimental Center for Archaeology	175	373

DISC- The Discovery Programme	71	616
FBK- Fondazione Brunno Kessler	98	517
POLIMI- Polytechnic of Milan	938	2050
MAP- Centre National de la Recherche Scientifique	626	625
MNIR- Muzeul National de Istorie a Romaniei	266	418
STARC- The Cyprus Institute	300	299
Archeotransfert	125	7155
CMC- CMC Associates	32	374
UJA - University Research Institute of Iberian Archaeology - University of Jaen	335	1003
CNR-ITABC - Consiglio Nazionale Delle Richerche	580	437
KMKG - Royal Museuems of Art and History	0	0
Total Number	**3.745**	**15.414**

Table 1: Statistics of Europeana

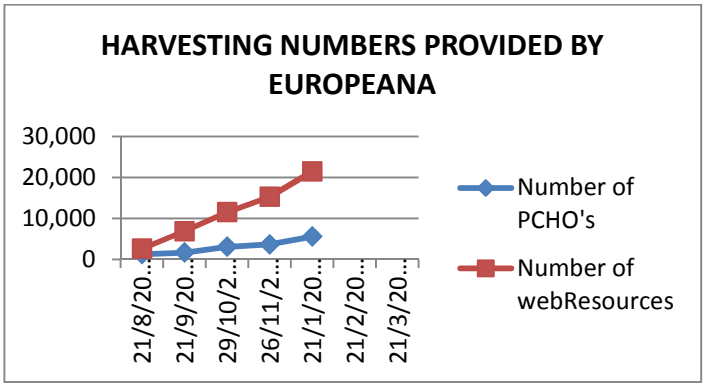
However, the following Table 2 shows statistics about the content that has been ingested to MORE2 and is ready for publication to Europeana as of 31/01/2015 (these figures include the items that have already been published).

Data provider name & acronym	**Number of PCHO's**	**Number of webResources**
CETI - Athena Research and Innovation Center	64	574
CNR-ISTI- Instituto di Scienza e Tecnologie dell Informazione	196	1.878
ICA- Interdepartimental Center for Archaeology	247	578
DISC- The Discovery Programme	376	1.170
FBK- Fondazione Brunno Kessler	115	551
POLIMI- Polytechnic of Milan	1.021	2.207
MAP- Centre National de la Recherche Scientifique	880	1.760
MNIR- Muzeul National de Istorie a Romaniei	379	606
STARC- The Cyprus Institute	871	868
Archeotransfert	127	7.202
CMC- CMC Associates	82	673

UJA - University Research Institute of Iberian Archaeology - University of Jaen	590	2.090
CNR-ITABC - Consiglio Nazionale Delle Richerche	156	591
KMKG - Royal Museuems of Art and History	457	16
Total Number	**5.561**	**20.764**

Table 2: Statistics of MORE2

The evolution of the content delivered and accepted for publication into EUROPEANA is shown in *Figure 8*



Date	Number of PCHO's	Number of webResources
21/8/2014	1.262	2.598
21/9/2014	1.677	6.865
29/10/2014	3.057	11.504
26/11/2014	3.662	15.255
21/1/2015 (*)	5.563	21.412

Figure 8. Timeline for total number of PCHO's and WebResources published in Europeana

*Figure taken from the Europeana development portal

5.4 3D-Icons Ingestion Statistics collected by the 3D-Icons portal

The 3D-Icons Portal provides a statistics page about content ingested to MORE2. In http://3dicons.ceti.gr/index.php/statistics , there is one table with full statistics per partner (Figure *9*) and one timeline graph (Figure *10*) that displays the progress of the harvesting. These figures displayed below.

Source	HA's	DR's	3D	Images	Videos	Other
1.ARCHEOTRANSFERT	127	7202	223	6956	23	0
2. CETI	64	574	170	393	11	0
3. CISA	247	578	120	450	8	0
4. CMC	82	676	41	618	17	0
5. CNR-ISTI	196	1966	167	1791	8	0
6.CNR-ITABC	128	567	195	372	0	0
7. CNRS-MAP	880	1760	498	1228	34	0
8. CYI-STARC	871	871	94	777	0	0
9. DISC	376	1262	193	995	74	0
10. FBK	115	552	71	461	20	0
11.KMKG	457	457	457	0	0	0
12. MNIR	379	606	198	408	0	0
13. POLIMI	1021	2207	1000	1176	31	0
14. UJA-CAAI	590	2090	593	1457	0	0
Total	5533	21368	4020	17122	226	0

Figure 9. Table Statistics per partner (23/1/2015)

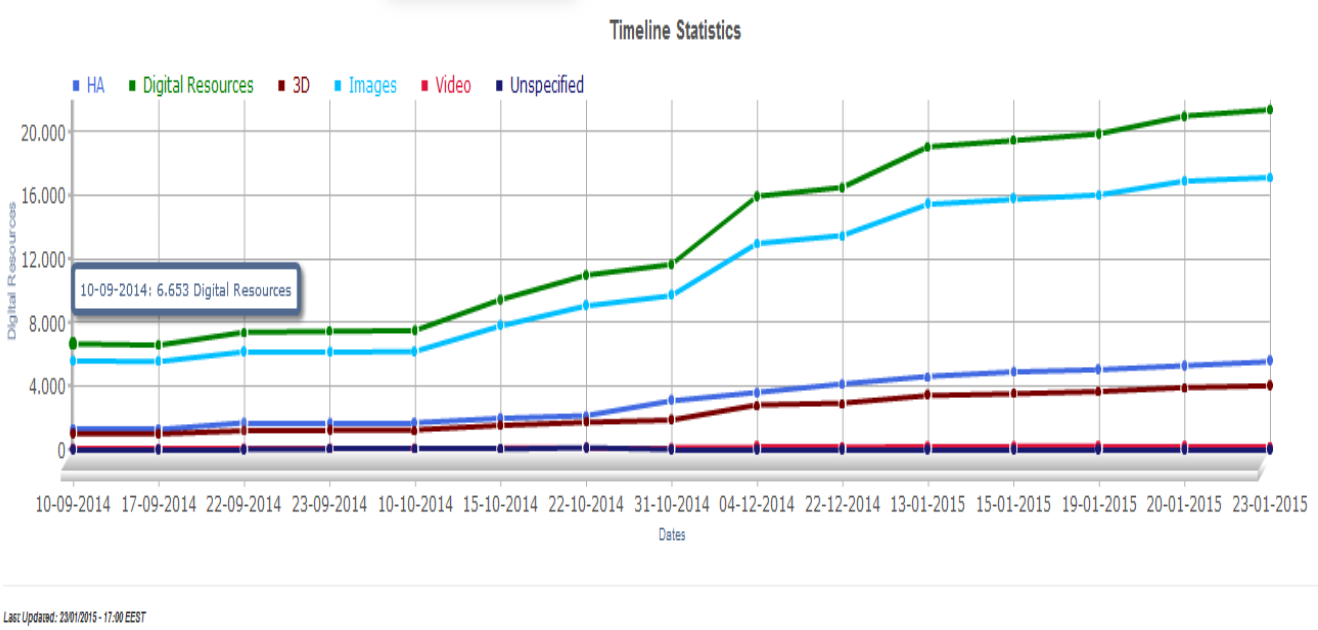


Figure 10. Timeline Statistics as recorded by the 3D-ICONS portal

In summary the status of the 3D-ICONS ingestion is shown in Table 3.

30/1/2015	3D-ICON's PORTAL						EUROPEANA (includes Harvesting of 23/1/2015)						
Data Provider Acronym	HA's	Digital Resources	3D's	Images	Videos	Europeana Collection Number	CHO's	Web Resources	3D Type CHO's (*)	Image Type CHO's (*)	Video Type CHO's (*)	Text Type CHO's	
CISA	247	578	120	450	8	2048703	248	587	121	119	8		
CNR-ITABC	156	595	223	372	0	2048714	156	595	150	6			(4)
CNR-ISTI	196	1966	166	1791	8	2048702	196	1966	169	24	3		
CETI	64	574	170	393	11	2048701	65	579	35	19	11		(1)
DISC	376	1262	193	995	74	2048705	376	1262	190	110	74	2	
UJA-CAAI	590	2090	590	1497	3	2048713	590	2090	590				(5)
CMC	82	676	41	618	17	2048712	82	676	41	26	15		(3)
POLIMI	1021	2207	1000	1176	31	2048707	1021	2209	503	498	20		
VisDim						2048716							
ARCHEOTRANSFERT	127	7202	223	6956	23	2048711	127	7203	107	9	9	2	
FBK	115	552	71	461	20	2048706	115	552	48	47	20		
KMKG	457	457	457	0	0	2048715	457	457	457				
CYI-STARC	871	871	94	777	0	2048710	871	871		777		94	(2)
CNRS-MAP	880	1760	498	1228	34	2048708	880	1760	249	614	17		
MNIR	379	606	198	408	0	2048709	379	605	99	280			
Total	5561	21396	4044	17122	229		5563	21412	2759	2529	177	98	

Table 3 Harvesting Status as of 30/1/2015

(*) These are CHOs. Each CHO contains more than one WR

(1) CETI There is a duplicate left over record from a previous harvesting. Thus the real numbers are 64,574,34,19,11

(2) STARC Their 94 3Ds were characterized as Monuments instead of 3D. Thus in Europeana are classified as TEXT but they are 3D

(3) CMC's thumbnails not harvested yet

(4) ITAB's thumbnails not harvested yet

(5) 3Ds, Images and Videos are all placed together. This is why we have no Image or Video CHO. The 3 Videos are in pdf files. Thus the portal counts them as 3D

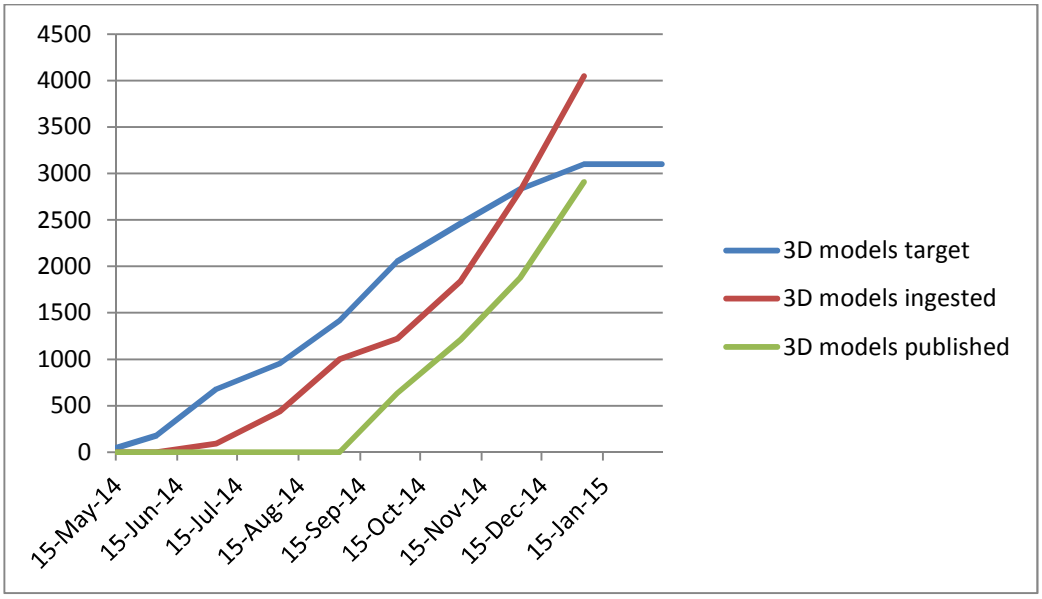
6. Ingestion Progress, Issues and Solutions

6.1 Ingestion Progress vs. Planned Schedule

Specific monthly ingestion targets were set for each partner from May 2014 to January 2015 for 3D models, images and videos. The totals per months, as presented in D4.2 Interim Report on Metadata Creation were as follows:

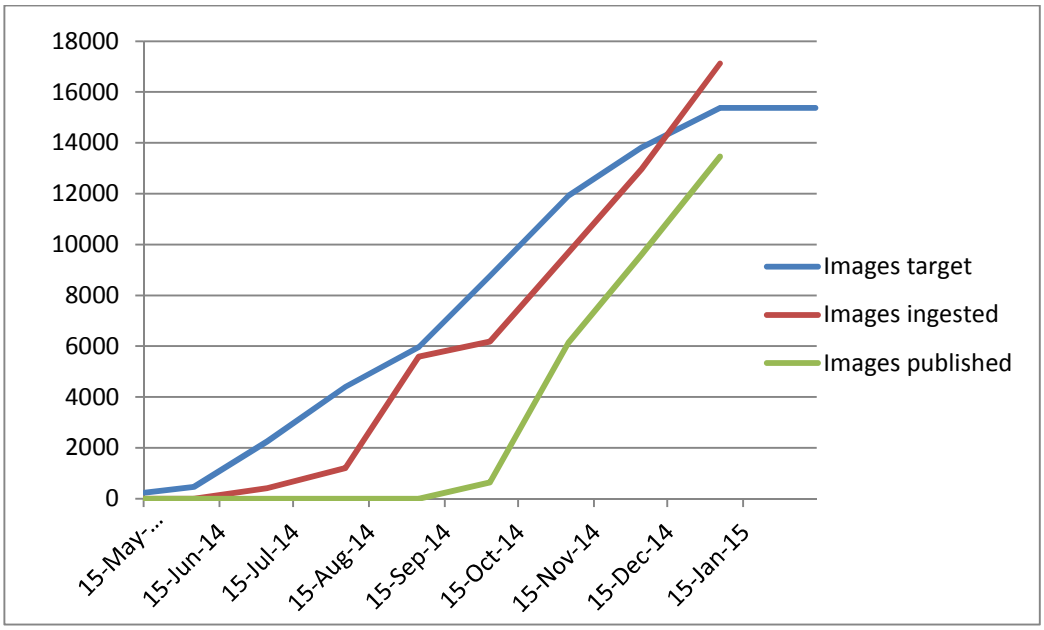
	Upload target date:									**Total**	**DOW**
	15-May-14	04-Jun-14	04-Jul-14	05-Aug-14	04-Sep-14	03-Oct-14	04-Nov-14	04-Dec-14	05-Jan-15		
3D	49	131	497	276	464	637	410	367	269	3,100	2,958
Images	234	229	1,781	2,156	1,561	2,798	3,147	1,907	1,566	15,379	13,191
Videos	2	7	31	79	40	58	44	22	11	294	166

The following graph shows the progress made with the ingestion and publication of 3D models.



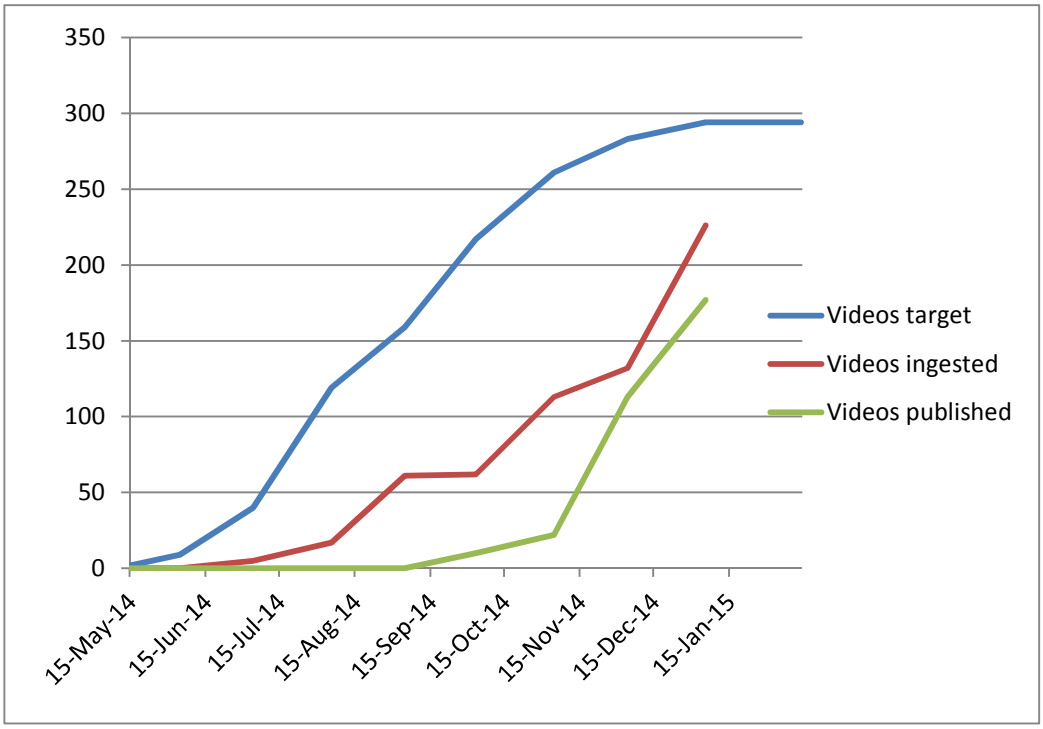
By January, there were over 2,900 3D models available through Europeana. Some of the content providers have used the carousel to group their 3D models. For example, Polimi have two 3D models per carousel, one low resolution for viewing and a high resolution dataset which can be supplied upon request. CETI has grouped models of different resolutions. The target was reached by November with just over 4,000 3D models in total being supplied to Europeana.

The following graph shows the progress made with the ingestion and publication of images.



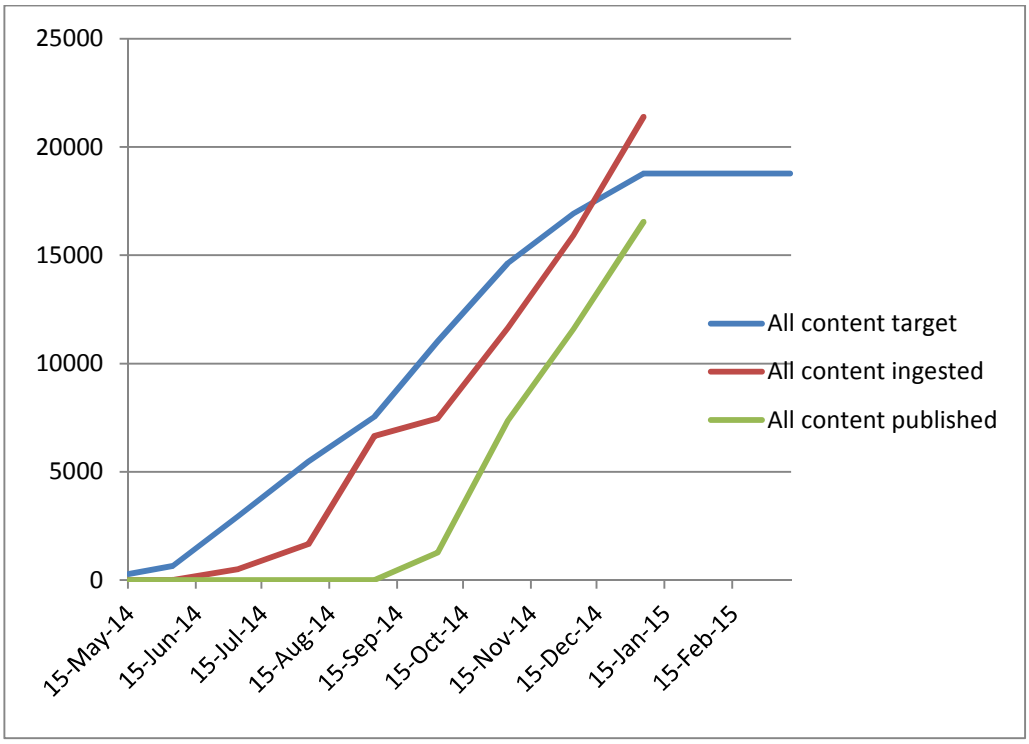
The number of images has also exceeded the target although the initial ingestion of images lagged behind as the partner's effort was concentrated on metadata for the 3D models. The carousel format has been used extensively for images in Europeana by 3D-ICONS.

The following graph shows the progress made with the ingestion and publication of videos.



This target (294) has not yet been reached but not all video metadata has been uploaded. The original figure in the Performance Monitoring Table was 100 videos.

The final table is for the combined content.



Overall, this total (18,773) has been exceeded and should be reached by the next publication by Europeana. Overall, it took four months before the first 3D-ICONS content was successfully published and will be nine months to achieve the targets.

6.2 Issues Encountered and Solutions Applied

The first metadata uploaded to MORE2 and the 3D-ICONS Portal was experimental and several issues were identified with it such as missing fields, non-working URLs and inconsistent rights statements. At this time, the majority of partners were using the Metadata Editor with UJA-CAAI, POLIMI and STARC opting to map existing metadata through MINT2. The first feedback came from Europeana on the 17th July which highlighted some additional issues which relate to Europeana requirements with respect to options available and how they render their metadata in their portal. One problem that Europeana has that information relating to options available to its content providers is dispersed and in some cases, non-existent. For example, around the beginning of 2014, Europeana amended the Rights Statements for Restricted Access, announcing this very fundamental change via the Blog which was consequently missed. For many partners, the original Rights Restricted statement suited their needs perfectly and the replacement options did not meet their needs. As a result of this, quite a few partners had to update their content Right statements and the Italian partners, due to the law relating to cultural heritage objects, had to use the (new) Paid Access statement provided by Europeana as this was the only option that met the national legislation despite no money being charged for access to their content. Another example that caused 3D-ICONS problems was the carousel format used by Europeana. As there is no

documentation for this, the initial ingestion was done based upon assumptions on how this worked; only by viewing the end results in the Europeana Portal and subsequent exchanges with the technical team were the partners able to understand how this was implemented and therefore how it could be used by the project effectively. Basically, because only one thumbnail (for the 1st Digital resource) is used and the carousel format was originally designed for series of images supplied by the MIMO project, thumbnails are created from the remaining image DR. However, if these digital resources are not of IMAGE type, then either the default icon is displayed or the thumbnail is empty. The project then issued Guidelines advising the partners to only use the carousel with images for multiple digital resources belonging to one PCHO (although the first DR can be of any type). However, this was not always possible to implement due to the nature of the data. Some partners are still using carousels with 3Ds and videos.

Other issues picked up by Europeana were duplicate URLs for the Landing pages (usually where partners had multiple related objects on one page), not everyone used the mandatory TYPE values in the Type field and several partners were initially confused by the difference between the IsShownAt and IsShownBy fields. Another common problem was that more than one name was used for an organization (so different names appeared in different records). Finally, the metadata supplied by UJA-CAAI and POLIMI revealed some issues in MINT2 (one of which was that the mapping hadn't been updated to the latest CARARE2 schema!) which NTUA fixed very quickly. Once this feedback was received, R.C. Athena (CETI and DCU) would investigate the problems to identify the underlying cause and then contact each partner concerned to advise them on which corrections had to be made to the metadata. Only one mapping was changed after the implementation of the CARARE2-EDM transformation and this was for the format field. Originally, the format field referred to the PCHO and what this was made from (e.g. marble or glass). However, since the PCHO (Heritage Asset) is the 3D model (rather than the physical object), then the format could refer also to the file format (e.g. 3D PDF, VRML). Both these options were mapped to the EDM:Format field which then led to some Format fields containing both physical material types and digital file types. Since the CARARE2 schema has other fields which specify the physical properties of the monument or object, the mapping was changed to just the file type since this is of more use to the Europeana end user (and consistent with the PCHO being a 3D model).

The second round of feedback from Europeana was sent in September by which time the situation had started to improve as the partners gained more insight and experience. The main problem reported was with the edm:Event class and the solution applied by Europeana was not to map this (this data is not displayed) and to proceed with publishing the rest of the metadata. In the meantime, the partners implemented technical solutions that enabled them to link and display to individual landing pages for each object which also linked through to a common aggregated content landing page. For example, CNR-ISTI display each object page which then redirects to the main content page for each PCHO which contains all related 3D models, images, videos and information. POLIMI implemented a solution which generates landing pages on the fly according to the DR selected. By November, the feedback picked up a few minor problems with some of the newest records and by

December, Europeana reported 0 “Blockers” (issues which prevent metadata being published) although some suggestions were made for improvements.

6.3 Ingestion of January 2015

The following three tables summarize the status of the project after the harvesting and ingestion to Europeana on January 21, 2015. The data for the 3Ds and the videos are taken from the experimental prepublication portal of EUROPEANA. The numbers of the images are taken from the 3D-Icons portal but they are in a good agreement with the total WebResources number provided by the Europeana Ingestion Team.

Participant number	Name	No. 3D Models in DoW	No. 3D Models in D4.2	Accepted in Europeana 31/01/2015
1	CISA	33	120	120
3	CNR-ISTI	69	155	166
	CNR-ITABC	116	185	223
4	CETI	30	72	170
5	DISC	85	124	193
6	UJA-CAAI	763	586	590
7	CMC	53	32	41
8	Polimi	527	527	1000
9	VisDim	0	50	
10	Archeotransfert	258	211	223
11	FBK	57	59	71
12	KMKG	450	455	457
13	CYI-STARC	71	71	94
14	CNRS-MAP	366	353	498
15	MNIR	80	100	198
	Total	2958	3100	4044

Table 4. 3Ds accepted by Europeana as of 31/1/2015

Participant number	Name	No. of Images in DoW Table	No. of Images in D4.2	Submitted to Europeana 21/01/2015
1	CISA	330	462	450
3	CNR-ISTI	795	860	1791
	CNR-ITABC		640	372
4	CETI	300	347	393
5	DISC	346	506	995
6	UJA-CAAI	1155	1461	1497
7	CMC	160	1074	618
8	Polimi	755	833	1176
9	VisDim		200	
10	Archeotransfert	6600	6600	6956
11	FBK	440	190	461
12	KMKG	0	0	0
13	CYI-STARC	510	512	777
14	CNRS-MAP	750	1194	1228
15	MNIR	550	500	408
	Total	12691	15379	17122

Table 5. Images accepted by Europeana as of 31/1/2015

Participant number	Name	No. Videos in DoW Table	No. Videos in D4.2	Accepted in Europeana 21/01/2015
1	CISA	0	8	8
3	CNR-ISTI	9	10	0
	CNR-ITABC		23	8
4	CETI	6	7	11
5	DISC	36	101	74
6	UJA-CAAI	5	3	3
7	CMC	4	18	17
8	Polimi	3	19	31
9	VisDim		10	
10	Archeotransfert	39	40	23
11	FBK	5	11	20
12	KMKG	20	20	0
13	CYI-STARC	13	13	0
14	CNRS-MAP	10	11	34
15	MNIR	2	0	0
	Total	152	294	229

Table 6 Videos accepted by Europeana as of 31/1/2015

6.4 Ingestion Planning for February and March 2015

Metadata harvesting and corrections will continue after the end of the 3D-Icons project (31st January 2015) to meet the required targets. Some partners have supplied all their agreed content but are happy to supply extra and to add some experimental content such as Sketchfab versions of High resolution 3D models. The current situation for each partner and the schedule is as follows:

Name	Pending (in Portal)	Total 3D records	Additional records to be uploaded after 31/01/2015	Status
CISA	0	120	0	Corrections to be made, ingestion February
CNR-ISTI	0	166	0	Completed, checking metadata
CNR-ITABC	58	223	0	Completed, checking metadata
CETI	0	170	0	Completed, checking metadata
DISC	22	212	0	Completed, checking metadata
UJA-CAAI	0	590	0	Completed, checking metadata
CMC	8	40	0	8 more 3D models to be uploaded (currently in MORE2)
Polimi	0	1000	20	Completed, checking metadata
VisDim	0	0	100	Entering metadata into the Editor, ingestion by March
Archeotransfert	21	223	0	Ingesting last few records, several corrections made
FBK	18	78	0	Completed, checking metadata
KMKG	0	457	0	Corrections to be made, ingestion February
CYI-STARC	74	94	50	Corrections to be made, ingestion February
CNRS-MAP	249	498	0	Corrections to be made, ingestion February
MNIR	101	200	0	Completed, checking metadata

Total	551	4071	170

The majority of partners have finished their metadata ingestion although FBK and POLIMI may add a few more models during February and March. CETI is waiting for approval to publish one more Byzantine Church. VisDim is currently entering their metadata and should complete by the March ingestion. In the meantime, everyone is checking their records in Europeana just to ensure that all the information is correct and complete and that the digital resources are present and usable.

7. Conclusions

3D is a relatively new media for presenting cultural heritage. Handling big data files, 3D web-viewers still under development, continuously evolving scanning process, (new technologies in scanning, more efficient scanning equipment, new algorithms in software), creating a metadata schema suitable for 3D cultural heritage objects, are some of the challenges the project had to face through its three years, creating several challenges in the initial stages of the project. Furthermore, as the majority of partners were new to Europeana, they did not anticipate the complexity of the Europeana metadata ingestion process or how much time this would take. However, overall the project met its promised targets., Table 4, Table 5 and Table 6 show the number of 3D models, images and videos as they were declared in the table of the DoW, the update of the DoW table as it was recorded in Deliverable 4.2 and the number of models that are published (currently in the prepublication experimental portal of Europeana) in Europeana as of 31/01/2015. Thus, for the 3D models, the 3D-ICONS project in the original DoW table stated 2,958 3Ds, then this number was revised in D4.2 to 3,100 and currently 3D-ICONS has uploaded 4,044 3D models. Regarding the other resources (images and videos), in the DoW the promised number was 12.691 Images and 152 Videos, in deliverable 4.2 these numbers were updated to 15,379 Images and 294 videos, while currently the project has published a total of 17,122 Images and 229 Videos. Some small discrepancies between the numbers reported by Europeana (higher) and the ones given in this Deliverable, see Table 3, are due to the fact that there are some duplicate records in Europeana. Most of these have already been identified and Europeana has been asked to remove them.

Annex I – Submission Information Package

The ingestion process utilizes a REST-based web service which is provided by the MORE2 repository. The service recognizes submission packages that are verified and ingested into the repository. Each submission package corresponds to a unique item. These submission packages have the following features:

File name:	[item_id].zip
File type:	Compressed file in zip format
Contents:	2 XML files • info.xml • carare.xml
info.xml	Contains the following information regarding an item: • content provider id *[mandatory]* • content provider name *[mandatory]* • user id (user who published the item) *[optional]* • user name (user who published the item) *[optional]* • native item identifier (unique) *[mandatory]* • native item name *[optional]*
info.xml	Contains the following information regarding a package: • package timestamp (created) *[mandatory]* • package size (total size of package) *[mandatory]* • items list (a listing of all items included in the package. Each item in the items list will contain the id, name attributes (as in the item level

	info.xml format) plus the filename of the item *[mandatory]*
native.xml	Contains the native record (well -formed xml)
carare.xml	Contains the carare record (well -formed xml)

The service can handle either **single package submissions** or **bundles of submission packages**. A submission package is a compressed (.zip format) file which contains multiple submission packages (as described above).

File hierarchy example of a bundle:

- upload_1.zip
    - info.xml
    - item_1.zip
        - info.xml
        - native.xml
        - carare.xml
    - item_2.zip
        - info.xml
        - native.xml
        - carare.xml
    - …

There are two different kinds of info.xml files: a) info.xml at the package level and b) info.xml at the item level. The info.xml files contain important information regarding either the package or the item. Details about the structure of the info.xml files and their contents are shown below:

Example of info.xml files [package level]:

```
<?xml version="1.0" encoding="UTF-8"?>
<package timestamp="" size="" >
<items>
        <item id="4" name="test 4" filename="item1" />
</items>
```

```
</package>
```

Example of info.xml files [item level]:

```
<?xml version="1.0" encoding="UTF-8"?>
<provider id=”2” name=”DCU” />
<user id=”13” name=”Dimitris Gavrilis” />
<item id=”51” name=”test” />
```

Annex II –Metadata Editor

The partners can choose from the following blocks. These are: Organization, Collection, Actor, Activity, Spatial data and Digital Resources (Figure 16).

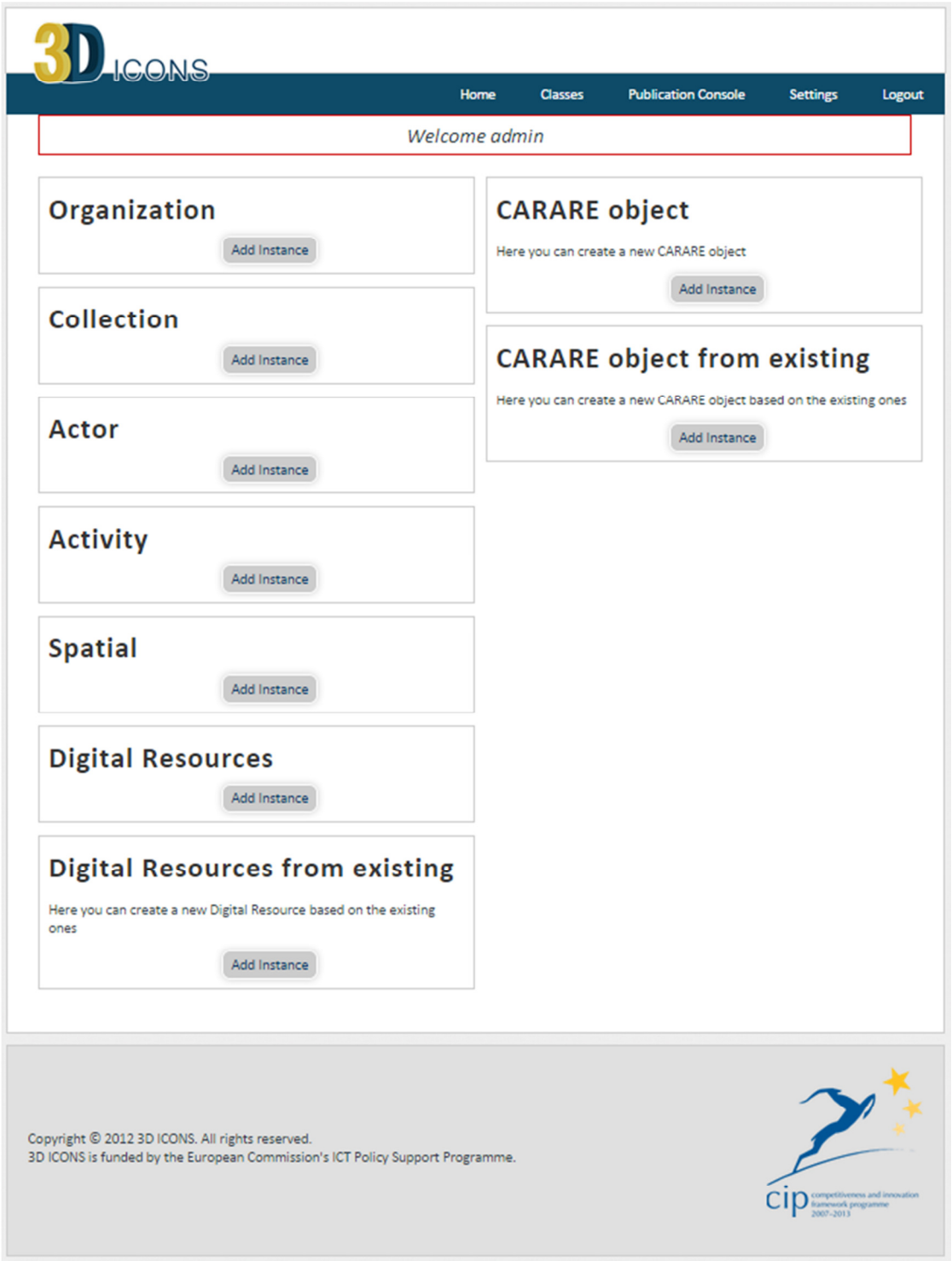


Figure 16. Home page of metadata editor:

Apart from the basic blocks presented above, providers are able to create Digital Resources or CARARE objects based on existing ones, due to the fact that many of them are much alike and differ only to a few elements.

Moreover, when a provider needs to correlate a Digital Resource to its Heritage Asset, he can choose from the already created Digital Resources grouped in categories (Figure 17).

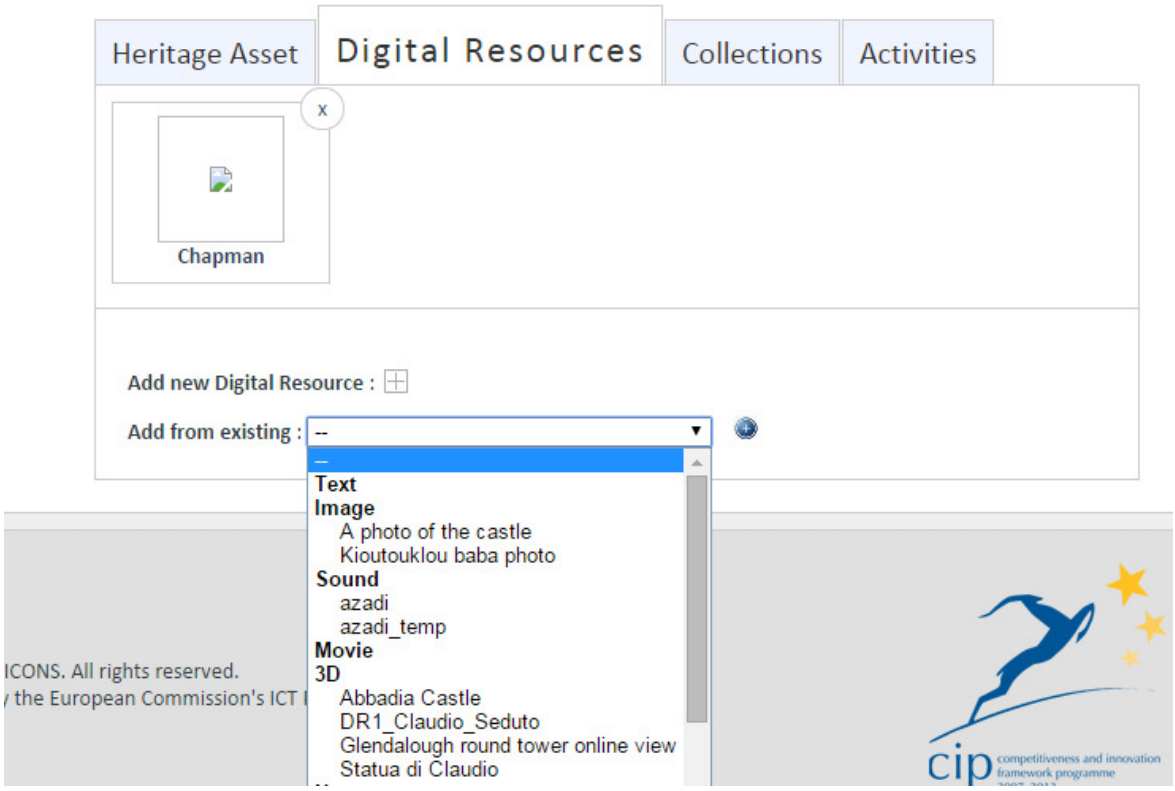
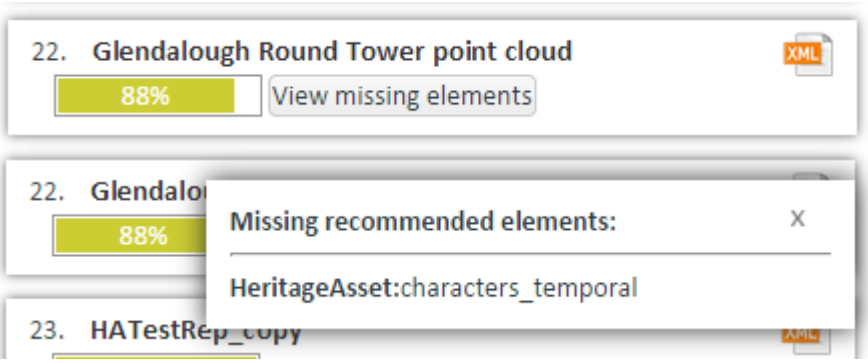


Figure 17. Correlate Digital Resources to Heritage Assets

Finally, providers are also able to view the elements missing from their Heritage Assets. This helps them to fill any missing mandatory or strongly recommended elements so as the record to be absolutely completed, be published to MORE2 and also be consistent with Europeana rules.



Annex III 3D-Icons portal: Presenting the data ingested in MORE2

DATA PRESENTATION

Geographic location is one of the most important attributes of every cultural heritage item. It can describe provenance, the current institution, the location of the event or other related events. The most valuable geographic description is in the form of digital geographic coordinates. Geographic coordinates presented as x, y define a position in a Cartesian coordinate system. The added value of the geocoded cultural content is in the browsing of cultural portals efficiently through space and time, searching for content in a more user friendly way, without the necessity of typing geographical names, making it possible to discover overlapping cultural content at the same location but originating from different sources and at different times, mapping the cultural content. The objective of the web application is to demonstrate the functionalities and advantages for displaying and browsing digital cultural content if a user interface is a map. The 3D-ICONS Portal (http://3dicons.ceti.gr) mapping web application consists of four main components:

- Map Engine
- Search Component
- Metadata Presentation
- Responsive Design

The pilot data will consist of all 3D models data ingested to Europeana and the geoparsed objects for the 3D-ICONS Project.

Functional specifications

- Mapping every object to geographical view
- Responsive design of the web application in order to be user friendly for different devices such as desktop PC, laptop, tablet, smartphones.
- Search component to find specific digital and physical objects. The search component implemented with checkboxes selection for increased usability. The search type is:
    - Search by Keyword
    - Search by Country
    - Search by Type of object
    - Search by Size of object
    - Search by Time Spans
- Smartphone usage of compass in order to display the nearest object according to user's current position.
- Display enriched information for every object according to the metadata content.
- Display country and town based object with mouse left click.

- For every object exist on the application there is a hyperlink direct to Europeana.
- Simple and friendly user interface.

Technical specifications:

- Use of Open Source Platform
- Usage of Europeana API
- Use of technological components used in Europeana ICT as much as possible
- Parsing and post processing of metadata

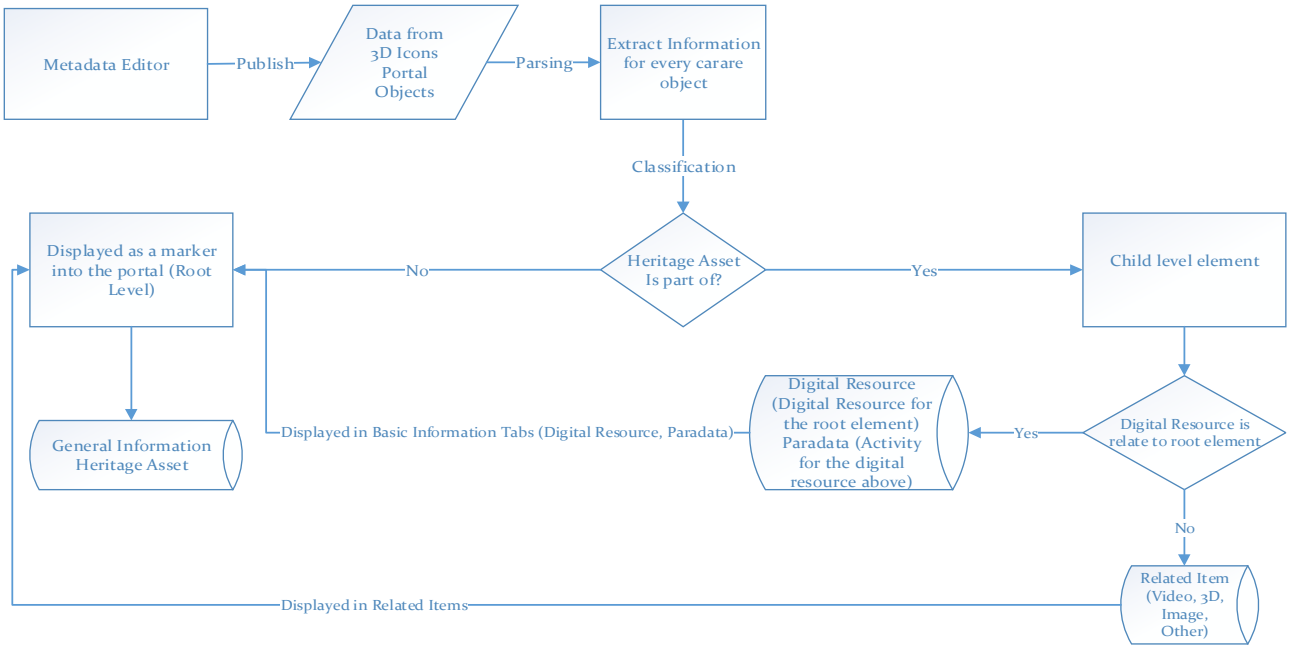


Figure 18. Architecture of the 3D-Icons Portal

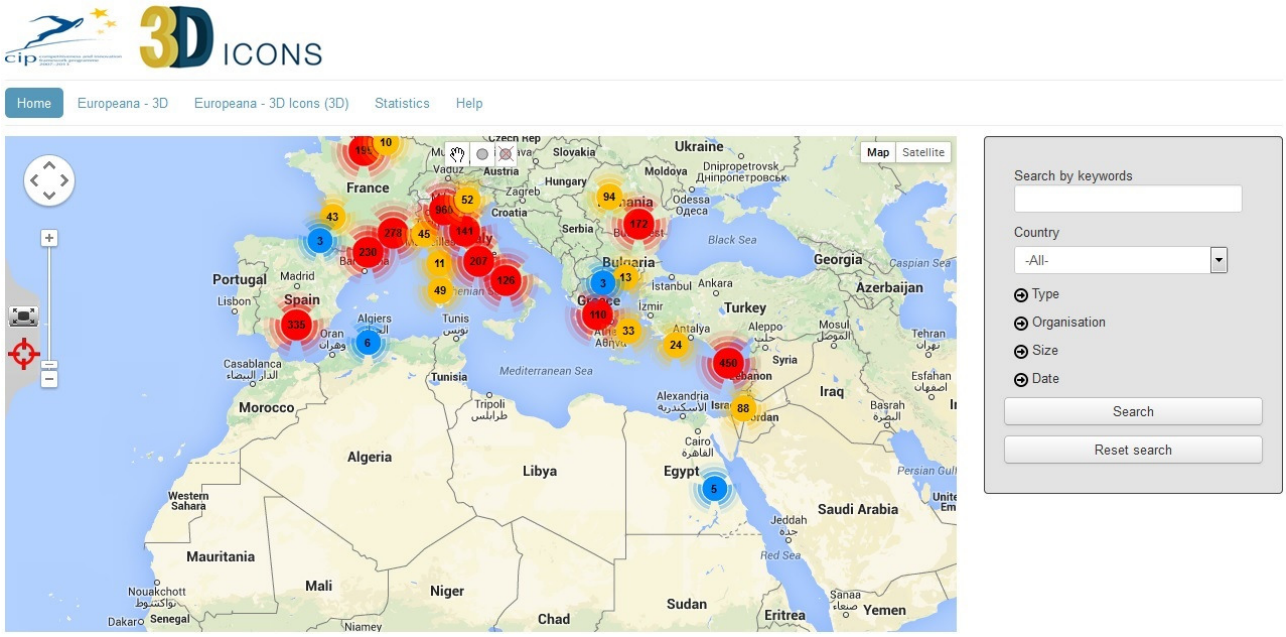


Figure 19. Home Page

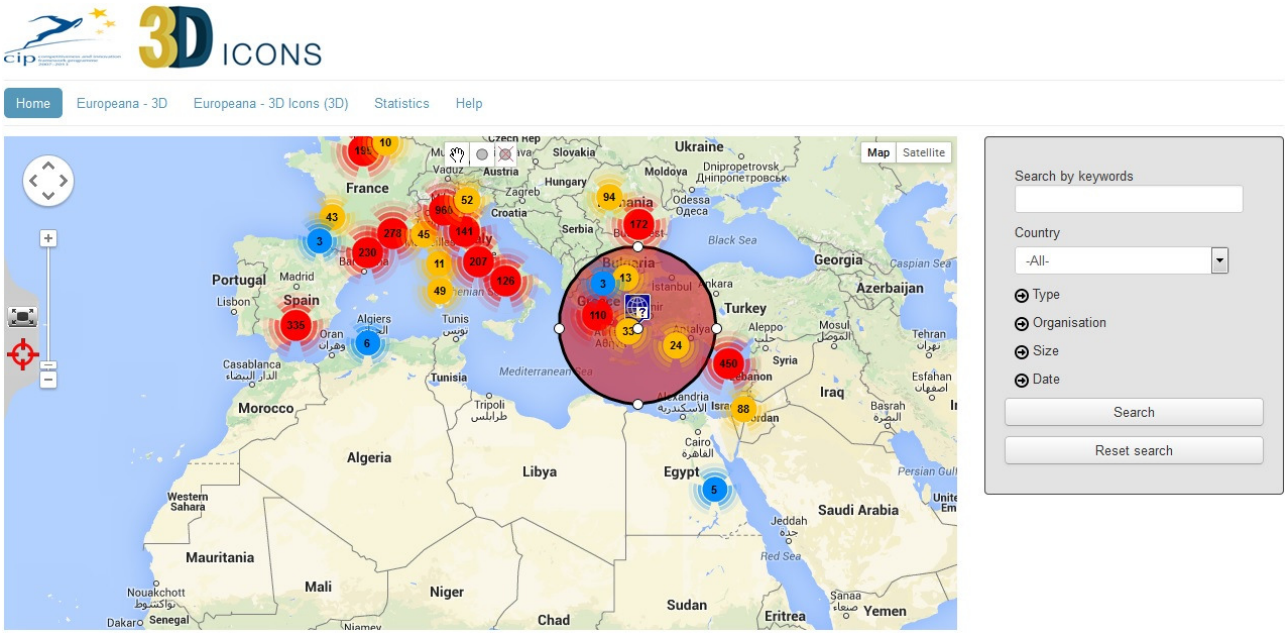


Figure 20. Search with radius / Search with Filters (Type, Organisation, Size, Date, Country)

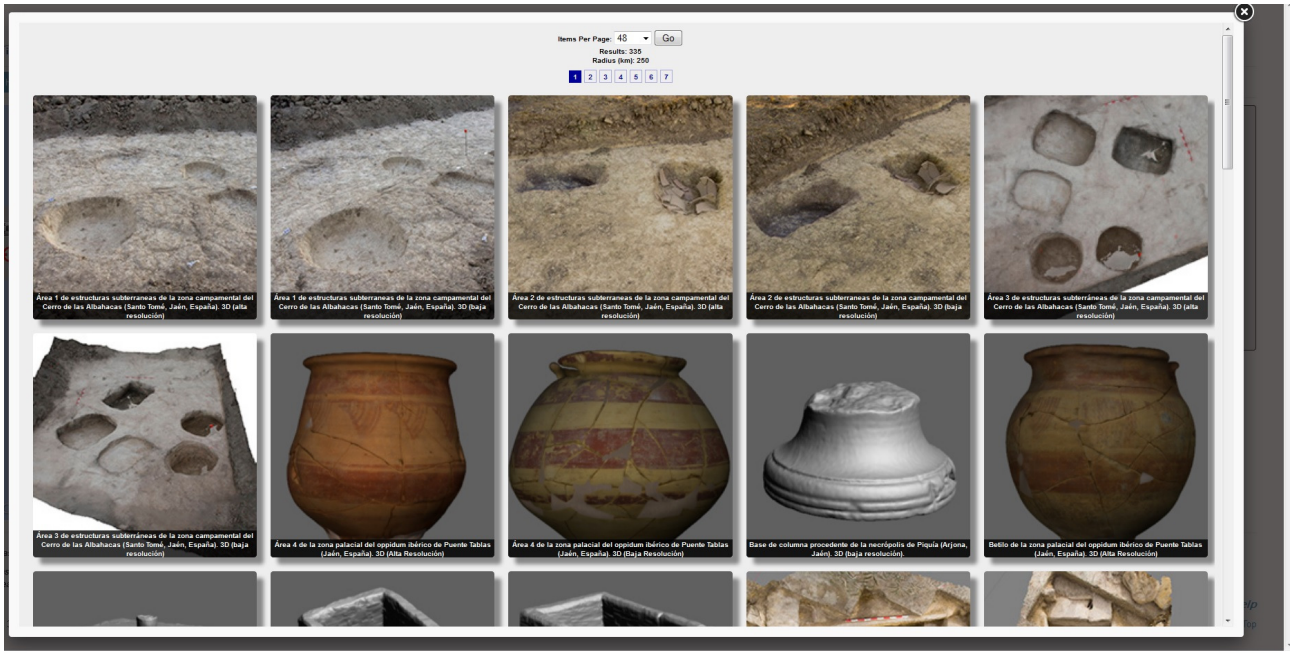


Figure 21. Display results

Figure 22. Display Record

Figure 23. Display Digital Resources images as gallery

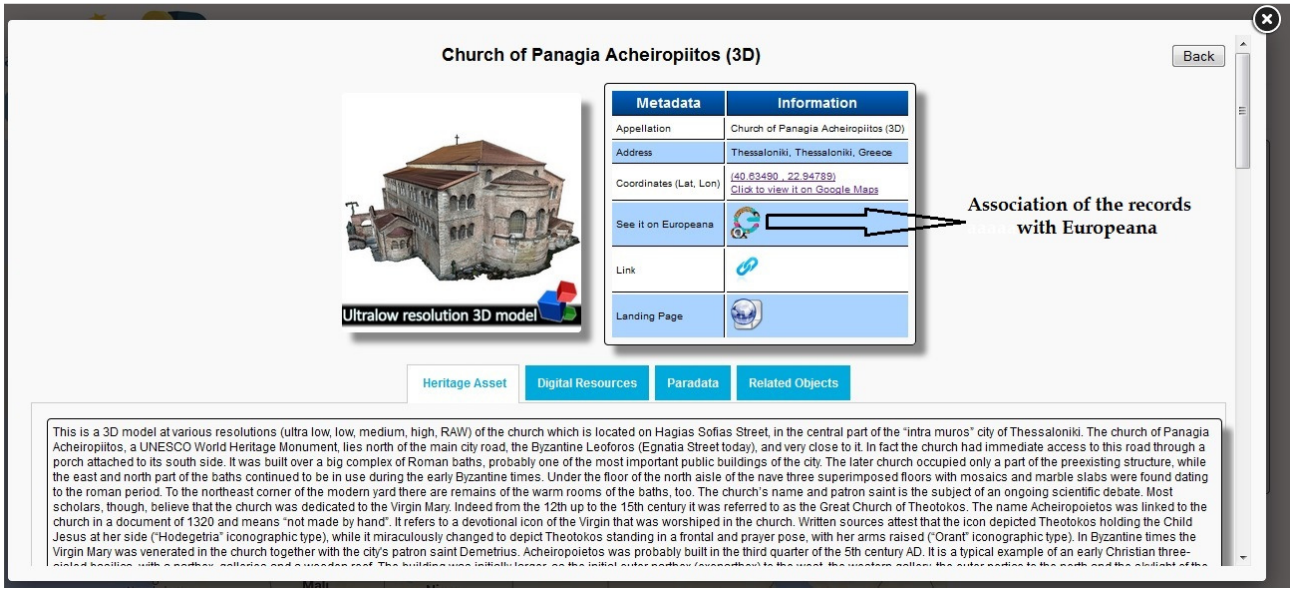


Figure 24. Association of the records with the uploaded to Europeana

References

D4.2 Interim Report on Metadata Creation