

INSIDDE

Collaborative Project

FP7 – 600849

D5.2.2 Report on the integration of 2D/3D digital models with movement into Europeana

Lead Author: Jose Antonio Sánchez

With contributions from: Marcos Fernández

Reviewer: Hamdi Dibeklioglu

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	December 2015
Actual delivery date:	December 2015
Version:	1.0
Total number of pages:	44
Keywords:	DSpace, Repository, Aggregator, Hispana, Regional Library of Varna, ESE, Europeana, 2D, 3D, INSIDDE

Abstract

This report documents the guidelines to follow for integrating the 2D and 3D models resulting from the project into Europeana. An essential requirement to contribute to the cultural heritage platform is to create and set up a repository that provides a mechanism to manage the digital surrogates and the metadata associated.

The INSIDDE project handles cultural objects from Bulgaria and Spain, therefore, the consortium had to collaborate with the aggregators from both countries that are the channel through which the metadata can be harvested and published on Europeana: Regional Library of Varna, the Bulgarian aggregator, and Hispana, the Spanish national aggregator.

Generally, aggregators harvest the metadata records from repositories by means of the OAI-PMH protocol. This makes necessary to select a repository that implements such protocol. The software repository selected is DSpace since it fulfils the requirements established and in addition the consortium is backed up by the GUDE, a Spanish DSpace users group, whose objective is to share experiences and knowledge about the implementation and support on DSpace.

Finally, the deliverable details the adaptation of the repository to meet the goal of integrating the digital objects into Europeana. The adaptation mainly consists of the installation of the Handle server for assigning, managing, and resolving persistent identifiers for the digital objects stored in the repository, a tool to generate the thumbnails of the objects and finally the plugin to export the metadata encoded in Dublin Core to the ESE format.

Executive summary

The deliverable D5.2.2 "Report on the integration of 2D/3D digital models with movement into Europeana" documents the guidelines to follow for integrating the 2D and 3D models resulting from the project into Europeana.

The first section deals with the Open Access (OA) philosophy. It explains what OA is, addresses the main goals of Open Archives Initiative (OAI) and describes the protocol for metadata harvesting Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

Following that, the second section focuses on defining repositories, their objectives, features and benefits. It also provides a typology proposed by different authors.

The third section gives an overview of the manner in which INSIDDE must interact with Europeana. The aggregators are also introduced. On one hand, the Regional Library of Varna – the Bulgarian aggregator – is in charge of harvesting the Bulgarian objects. Their harvesting method is simple; they require a single eXtensible Markup Language (XML) file in European Semantic Elements (ESE) format for every digital object provided. On the other hand, Hispana – the Spanish national aggregator – is in charge of harvesting the Spanish objects. They use the OAI-PMH protocol in order to harvest the metadata records.

The next section documents the study of the three main free and open license repository platforms in the market: Fedora, DSpace and Eprints.

The software repository selected is DSpace since it fulfils the requirements established, it was recommended by Hispana and in addition has the Grupo de Usuarios de DSpace de España (GUDE), a Spanish DSpace users group, whose objective is to share experiences and knowledge about the implementation and support on DSpace.

The fifth section details the adaptation of the repository to meet the goal of integrating the digital objects into Europeana. The adaptation mainly consists of the installation of the Handle server for assigning, managing and resolving persistent identifiers for the digital objects stored in the repository, development of a tool to generate the thumbnails of the objects and finally the implementation of the plugin to export the metadata encoded in Dublin Core to the ESE format.

The sixth section presents the content of the INSIDDE repository, which can be accessed through <http://idipre.treelogic.com/xmlui/>, and consists of two collections: one housing the Bulgarian objects and the other housing the Spanish objects.

The Bulgarian collection contains the 3D models of four vessels property of RIM Stara Zagora. In turn, the Spanish collection contains 2D images of 15 paintings belonging to the Museum of Fine Arts of Asturias plus one image resulting from an infrared analysis carried out by CNR-INO. The metadata was harvested by both aggregators in the first half of October. According to the aggregators, the content is planned to be published in Europeana in December 2015. This section comes before the conclusions section coming out from the analysis of the content of this deliverable.

Document Information

IST Project Number	FP7 - 600849	Acronym	INSIDDE
Full Title	Integration of technological solutions for imaging, detection, and digitisation of hidden elements in artworks		
Project URL	http://www.insidde-fp7.eu/		
Document URL			
EU Project Officer	José María del Águila Gómez		

Deliverable	Number	D5.2.2	Title	Report on the integration of 2D/3D digital models with movement into Europeana
Work Package	Number	WP5	Title	Integration of enriched surrogates into Europeana and development of an end-user application

Date of Delivery	Contractual	M36	Actual	M36
Status	version 1.0		final <input checked="" type="checkbox"/>	
Nature	prototype <input type="checkbox"/> report <input checked="" type="checkbox"/> demonstrator <input type="checkbox"/> other <input type="checkbox"/>			
Dissemination level	public <input checked="" type="checkbox"/> restricted <input type="checkbox"/>			

Authors (Partner)	Treelogic			
Responsible Author	Name	Jose Antonio Sánchez	E-mail	joseantonio.sanchez@treelogic.com
	Partner	Treelogic	Phone	00 34 985 966 136

Abstract (for dissemination)	<p>This report documents the guidelines to follow for integrating the 2D and 3D models resulting from the project into Europeana. An essential requirement to contribute to the cultural heritage platform is to create and set up a repository that provides a mechanism to manage the digital surrogates and the metadata associated.</p> <p>The INSIDDE project handles cultural objects from Bulgaria and Spain, therefore, the consortium had to collaborate with the aggregators from both countries that are the channel through which the metadata can be harvested and published on Europeana: Regional Library of Varna, the Bulgarian aggregator, and Hispana, the Spanish national aggregator.</p> <p>Generally, aggregators harvest the metadata records from repositories by means of the OAI-PMH protocol. This makes necessary to select a repository that implements such protocol. The software repository selected is DSpace since it fulfils the requirements established and in addition the consortium is backed up by the GUDE, a Spanish DSpace users group, whose objective is to share experiences and knowledge about the implementation and support on DSpace.</p>
-------------------------------------	--

	Finally, the deliverable details the adaptation of the repository to meet the goal of integrating the digital objects into Europeana. The adaptation mainly consists of the installation of the Handle server for assigning, managing, and resolving persistent identifiers for the digital objects stored in the repository, development of a tool to generate the thumbnails of the objects and finally the plugin to export the metadata encoded in Dublin Core to the ESE format.
Keywords	DSpace, Repository, Aggregator, Hispana, Regional Library of Varna, ESE, Europeana, 2D, 3D

Version Log			
Issue Date	Rev. No.	Author	Change
15/05/2015	0.1	Jose Antonio Sánchez	State of art of repositories
19/05/2015	0.2	Jose Antonio Sánchez	Repositories definition, etc.
27/05/2015	0.3	Jose Antonio Sánchez	OA info
8/06/2015	0.4	Jose Antonio Sánchez	Hispana
20/08/2015	0.5	Jose Antonio Sánchez	Several Content (DSpace)
27/10/2015	0.6	Jose Antonio Sánchez	Regional Library of Varna
23/11/2015	0.7	Jose Antonio Sánchez	Several Content (DSpace)
25/11/2015	0.8	Jose Antonio Sánchez	Several Content added
27/11/2015	0.9	Jose Antonio Sánchez	Minor changes
30/11/2015	0.10	Marcos Fernández Díaz	Minor changes
04/12/2015	0.11	Javier Gutiérrez Meana	Minor changes
13/12/2015	0.12	Hamdi Dibeklioglu	Revision
17/12/2015	1.0	Javier Gutiérrez Meana	Final version

Table of Contents

Executive summary.....	5
Document Information.....	7
Table of Contents	9
List of figures	11
List of tables.....	13
Abbreviations.....	15
1 Introduction to OA	17
1.1 OAI.....	17
1.2 OAI-PMH	17
2 Repositories: Definition and typology.....	19
2.1 What is a repository?	19
2.2 Typology of repositories	20
3 Aggregators, Europeana and metadata harvesting	23
3.1 Hispana.....	24
3.2 Regional Library of Varna.....	25
4 State of the art of repositories.....	26
4.1 Fedora	26
4.2 DSpace.....	26
4.3 EPrints	27
4.4 Election of repository software.....	28
5 The INSIDDE repository	29
5.1 Installation of DSpace	29
5.1.1 Software pre-requisites	29
5.2 Technical specifications	29
5.3 Customising Dspace	30
5.3.1 Persistent identifiers of digital objects	30
5.3.2 Thumbnails in DSpace.....	31
5.3.3 Plugin Crosswalk Dublin Core to ESE	33
5.4 OAI validation.....	34
6 Content and Metadata Harvesting.....	36
7 Conclusions	39
References.....	41
Annex A Regional Library of Varna ESE template	43

List of figures

Figure 1: Basic functioning of OAI-PMH.	18
Figure 2: Scheme of interaction with Europeana.	23
Figure 3: INSIDDE repository in Hispana.	25
Figure 4: OAI-PMH validation.	34
Figure 5: OAI-PMH validation result.	35
Figure 6: INSIDDE repository (Bulgarian collection).	36
Figure 7: INSIDDE repository (Spanish collection).	37

List of tables

Table 1: Hardware specifications.	29
Table 2: Software specifications.....	30

Abbreviations

2D: 2-Dimensional

3D: 3-Dimensional

API: Application Programming Interface

AVI: Audio Video Interleave

BMP: Bitmap

BOAI: Budapest Open Access Initiative

CNRI: Corporation for National Research Initiatives

CPU: Central Processing Unit

EDM: Europeana Data Model

ESE: Europeana Semantic Elements

GB: Gigabyte

GIF: Graphics Interchange Format

GUDE: Grupo de Usuarios de DSpace de España

HDD: Hard Disk Drive

HTML: HyperText Markup Language

HTTP: HyperText Transfer Protocol

ID: Identifier

IETF: Internet Engineering Task Force

JPEG: Joint Photographic Experts Group

MIT: Massachusetts Institute of Technology

MP3: MPEG Audio Layer III

OA: Open Access

OAI: Open Archives Initiative

OAI-ORE: Open Archives Initiative Object Reuse and Exchange

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting

OS: Operating System

PDF: Portable Document Format

PNG: Portable Network Graphics

RAE: Research Assessment Exercise

RDF: Resource Description Framework

RFC: Request for Comments

RSS: Really Simple Syndication

R&D: Research and Development

SDK: Software Development Kit

SRU: Search/Retrieve via URL

SWORD: Simple Web-service Offering Repository Deposit

TCP: Transmission Control Protocol

TIFF: Tagged Image File Format

UDP: User Datagram Protocol

URN: Uniform Resource Name

URL: Uniform Resource Locator

XML: eXtensible Markup Language

1 Introduction to OA

The term Open Access (OA) can be defined based on the Budapest Open Access Initiative (BOAI) declaration [1], signed in 2001, which states the following:

“By “open access” to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.”

Researchers, institutions, nations and society as a whole benefit from Open Access. For researchers, it brings increased visibility, usage and impact for their work. Institutions enjoy the same benefits in aggregated form. There is growing evidence to show that countries also benefit because OA increases the impact of the research in which they invest public money and therefore there is a better return of investment. Society as a whole benefits because research is more efficient and more effective, delivering better and faster outcomes for everyone.

There are two ways in which authors can provide open access to their work. One way is to publish it and then self-archive it in a repository. This is known as 'green' open access. The second way authors can make their work open access is by publishing it (in a journal or book) in such a way that makes their research output immediately available from the publisher. This is known as 'gold' open access.

1.1 OAI

The Open Access Initiative (OAI) develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content (metadata). The OAI has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. It attempts to build a "low-barrier interoperability framework" for archives (institutional repositories) containing digital content (digital libraries). It allows people (service providers) to harvest metadata (from data providers). This metadata is used to provide "value-added services", often by combining different data sets.

OAI has been involved in developing a technological framework and interoperability standards for enhancing access to e-print archives, which make scholarly communications like academic journals available, associated with the open access publishing movement. The relevant technology and standards are applicable beyond scholarly publishing.

1.2 OAI-PMH

The Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH) [6] protocol is a protocol developed by the OAI. It offers a simple technical option for catalogue and repository services to make their metadata available to other services, based on the Hypertext Transfer Protocol (HTTP) and Extensible Markup Language (XML) standards, resulting in facilitated discovery of distributed resources. The metadata to be harvested may be in any agreed format, although unqualified Dublin Core is required by the specification in order to provide a basic level interoperability.

The OAI framework distinguishes between data providers and service providers. Data providers have metadata that they wish to make available, typically the metadata will come from a catalogue or repository. Service providers create value-added services, such as search interfaces, based on the metadata from one or more data provider. The OAI-PMH allows data providers to make their metadata available for 'harvesting' by service providers; this is a process of collecting, as the metadata is gathered from a number of distributed repositories into a combined data store.

The aim of the OAI-PMH is to make easy the resource discovery process when resources are stored in several and independent repositories by exporting metadata about items in those repositories. The OAI-PMH protocol can be applied to a wide range of digital objects, such as images, 3D models, technical reports, etc. In order to implement the broadest level of interoperability, OAI-PMH requires the metadata

to be presented in Dublin Core. However the protocol accepts multiple forms of metadata to be exposed and any information associated with a resource that can be encoded in XML may be exchanged using OAI-PMH. Although the OAI-PMH is associated with the open sharing public metadata, it can be used for private exchange in closed systems.

OAI-PMH is based on client-server architecture (see Figure 1). The 'harvesters' request information of updated records from repositories. A harvester plays the role of a client application that performs OAI-PMH requests and collects ('harvests') metadata from repositories. A repository is a server accessible via the Internet network that implements the six OAI-PMH requests/verbs described below, and is managed by a data provider to expose metadata to harvesters. Data providers manage the repository and publish the resources and make the metadata available for harvesting by service providers. The OAI-PMH distinguishes between three distinct entities related to the metadata made accessible by the OAI-PMH: a resource (what the metadata is about); an item (a constituent of a repository from which metadata about a resource can be disseminated); and a record (metadata expressed in a single format). OAI-PMH specifies that unique identifiers are provided for items.

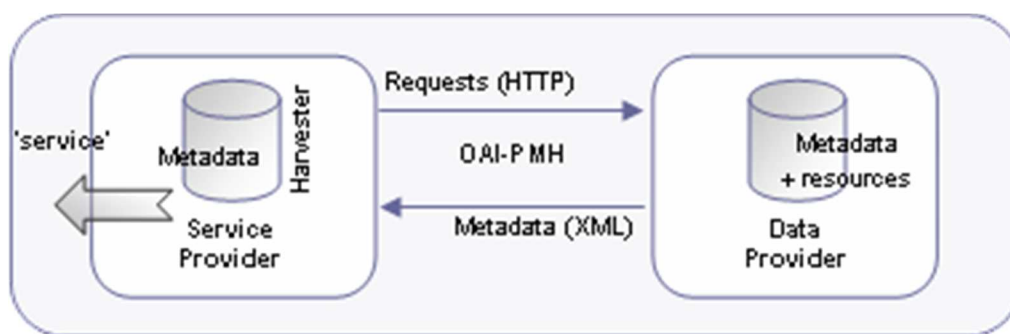


Figure 1: Basic functioning of OAI-PMH.

The OAI-PMH defines the following lists of requests or verbs:

- *Identify*: This verb is used to retrieve information about a repository. Some of the information returned is required as part of the OAI-PMH.
- *ListMetadataFormats*: This verb is used to retrieve the metadata formats available from a repository.
- *ListSets*: This verb is used to retrieve the set structure of a repository, useful for selective harvesting.
- *ListRecords*: This verb is used to harvest records from a repository.
- *ListIdentifiers*: This verb is an abbreviated form of ListRecords, retrieving only headers rather than records.
- *GetRecord*: This verb is used to retrieve an individual metadata record from a repository.

2 Repositories: Definition and typology

An essential requirement to contribute to the cultural heritage platform Europeana is to create and set up a repository that provides a mechanism to manage the digital surrogates and the metadata associated. The following sections focus on defining repositories, their objectives, features and benefits. A typology proposed by different authors is also provided.

2.1 What is a repository?

A digital repository can be defined as an archive where digital resources (files, images, audio, etc.) can be stored. They arose from the so-called e-print community concerned about disseminating their scientific papers.

Another way of defining digital repositories can be found in [3]. It states that, in the short term, a digital repository:

- Enables quick, easy, simultaneous and remote access to deposits.
- Allows institutions/organisations to efficiently retain and manage their own intellectual assets. Digital repositories are being increasingly seen as a valuable tool for the Research Assessment Exercise (RAE).
- Facilitates the reuse of deposited materials for new research, education, and learning.
- Minimises physical storage requirements whilst increasing the potential mass of deposits.
- Manages both metadata and intellectual object in the same location.
- Enables external validation of research results.

Over the longer term, a digital repository:

- Can enable persistent access to deposits independently of external publishers.
- Can be used to store incremental deposits of unique observational data to spatially significant and new collections for developmental analysis.
- Increases institutional research visibility.
- Raises potential return on investment from asset creation.
- Enables long-term proof of authorship or assurance of credibility for unpublished papers if repository is certified/trusted.

A digital repository is a centralised platform created by an organisation in order to create, store and manage digital information, facilitating the users' searches and also the retrieval and access. Digital repositories must be developed based on metadata standards and protocols for collecting and exchanging information, implementing mechanisms to import, export and also collect the digital contents and metadata.

Repositories are typically implemented by means of open source software and applications and interoperability is a technical characteristic of digital repositories. "The OAI [4] develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. OAI has its roots in the open access and institutional repository movements. Continued support of this work remains a cornerstone of the Open Archives program. Over time, however, the work of OAI has expanded to promote broad access to digital resources for eScholarship, eLearning, and eScience".

The OAI empowers the creation of open and distributed repositories containing, at least, metadata describing the digital resources. Its purposes are to create and promote interoperability standards contributing to facilitate the efficient dissemination of the contents of the repositories.

2.2 Typology of repositories

There are many approaches to classifying the different types of repositories, some of them agree and others are based on analysis from different angles. Most of the authors agree on two types of repositories: disciplinary and institutional repositories.

Disciplinary repositories are archives where the outcomes, studies or material of research in a particular subject of study are made accessible. Any researcher in the field is able to include content. Disciplinary repositories provide access to journal articles (pre- or post-print), but may also provide access to other types of information and services.

Most of these repositories provide tools for metadata searching and links to the full text. Sometimes the full text is stored in the repository itself.

Disciplinary repositories are not uniform in nature. Some disciplinary repositories accept additional types of content, such as datasets, presentations, and working papers. Some offer additional services, like Really Simple Syndication (RSS) feeds, collaboration tools, and bibliography and curriculum vitae services.

Most of the time, disciplinary repositories lack of a formal organisation, there is not a formal structure, the repository itself is the organisation. They can acquire their content in many ways. Many rely on author or organisation submissions, others crawl the web for scholar and researcher websites and download publicly available academic papers from those sites, and others grew as a result of active involvement of academia and societies. A disciplinary repository generally covers one broad based discipline, with contributors from many different institutions supported by a variety of funders; the repositories themselves are likely to be funded from one or more sources within the subject community. Deposit of material in a disciplinary repository is sometimes mandated by research funders.

Institutional repositories are a common digital content management tool of an institution to support learning and research as well as a storage location and preservation. Institutional repositories are also often seen as the seal of quality and distinguishing mark of the institution with regards to the research authors who are part of it, promoting the visibility and dissemination of the scientific production of the institution.

Clifford A. Lynch defines institutional repository in [5] as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organisational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organisation and access or distribution".

Other authors, like Chris Armbruster and Laurent Romary establish in [6] an analysis of the existing typology of repositories, after two decades of implementation. They present an ideal classification of four types of repositories:

- Subject-based repositories (commercial and non-commercial, single and federated). Usually they have been set up by community members and are adopted by the wider community. Spontaneous self-archiving is prevalent as the repository is of intrinsic value to scholars. Authors have the opportunity to communicate ideas and results early in the form of working papers and preprints, from which a variety of benefits may result, such as being able to claim priority, testing the value of an idea or result, improving a publication prior to submission. They are well defined and alert services and usage statistics are meaningful for community users.
- Research repositories. Most of them are sponsored by public funding or private R&D investment in order to make the research results accessible. This capturing typically requires a mandate. Publications are results, including books, but data may also be considered a result worth capturing, leading to a collection with a variety of items. These repositories house rigorous results due to its content are peer-reviewed multiple times and the production of the results is well funded.
- National repository. They are systems that require coordination (more for a federated system, less for a unified system). National systems are designed to capture scholarly output more generally and not just with a view to preserving a record of scholarship, but also to support, for example,

teaching and learning in higher education. It serves scholarly communication in the national language and supports public policy, e.g. in generating open educational resources for higher education and enhancing public access to knowledge.

- Institutional repositories. They contain a variety of outputs of the Institution. It is a library holding an institutional collection, and it is a showcase because the online open access display and availability of the collection may serve to impress and connect with the members of the institution. An institutional repository may also be an instrument of the institution by supporting internal and external assessment as well as strategic planning.

3 Aggregators, Europeana and metadata harvesting

At the beginning of this task in July 2014 (M19), Treelogic (already a member of Europeana) contacted Europeana Professional [8] – “where librarians, curators and archivists share digital expertise” – with the purpose of being informed about the requisites to act as an aggregator. Europeana’s Aggregation team communicated that a partner institution interested in contributing data to Europeana would need to fill in the Data Exchange Information Form. This form identifies types and numbers of objects, data owners and other additional information regarding the digital resources to be shared.

Once filled in and returned back, they introduced the consortium to Hispana [9], the National Aggregator from Spain and the channel through which the metadata can be harvested and published on Europeana.

In order to clarify the manner in which INSIDDE must interact with Europeana, a scheme is presented in Figure 2. Understanding how the servers are interconnected is essential to upload content to the network and, therefore, to Treelogic, Regionalen Istoricheski Muzei Stara Zagora, and Museum of Fine Arts of Asturias. From our side as data providers, we are responsible for:

- The generation and storage of metadata: Metadata is generated taking into account the specific characteristics of the artwork and the resulting scans. It will be stored in a repository managed by Treelogic which implements the OAI-PMH protocol. As can be seen in [7], a link to the website where the digital material can be visualised is mandatory as part of this information.
- The communication with the aggregator: Aggregators in the context of Europeana are certified entities that act as gateways in the sense that they “collect metadata from a group of data providers and transmit it to Europeana”. Therefore, a communication between the INSIDDE repository which implements OAI-PMH protocol and the aggregator OAI-PMH repository must be established. As can be seen in following sections, the metadata harvesting by the aggregators is not always done using the OAI-PMH protocol, some of them just require an XML file with the metadata in the proper schema.

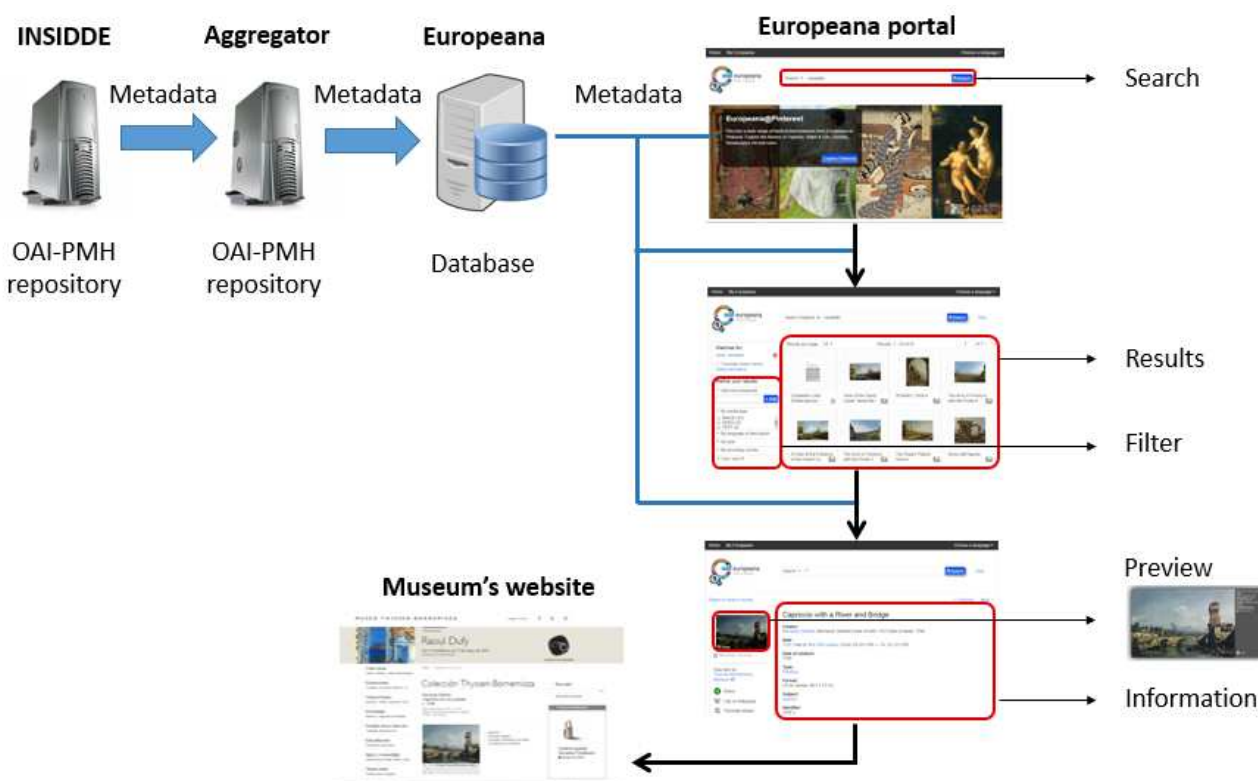


Figure 2: Scheme of interaction with Europeana.

Although it may seem to be complex, this process is transparent to the users: when they enter a query, the search engine employs the metadata stored in the database to produce the results. Likewise, to display the main characteristics of the artwork on the screen (including a link to the digital surrogate), metadata is analysed and processed. An icon allows the user to go directly to the repository for additional details.

The following sections present the two aggregators with which the INSIDDE consortium has worked in order to upload the outcomes of the project to Europeana.

3.1 Hispana

Hispana is a directory and harvester of digital resources and the common access point to digital objects of Spanish Memory Institutions, mainly libraries, museums and archives, and plays, in relation to Spanish digital repositories, a similar role to Europeana regarding European repositories. It is therefore an aggregator of content which complies with the OAI promoted by the European Union.

Spanish institutions interested in sharing their collections through Europeana do so in collaboration with the national aggregator. Both public administrations and non-profit organisations – e.g. foundations – are currently providing data to Hispana.

Hispana is heir to the Directory and Harvester of Digital Resources. It began in September 2005 as an initiative of the Library Coordination Unit of the Ministry of Education, Culture and Sport. Its objective was to provide a tool to allow the coordination of digitisation projects carried out in Spain to avoid repeating the scanning of the same work. It was released in March 2006 with 55 digitisation projects and 116 014 digital objects. In 2009, it was officially launched under the name of Hispana, with more than 1 million digital objects and became the Spanish national cross-domain aggregator for Europeana. It now supports Spanish libraries, archives, museums and other institutions to make their collections available via Europeana.

This initiative follows the recommendation of the European Commission on digitisation and online accessibility of cultural material and digital preservation on August 24th 2006, which aims to contribute to a coordinated action carried out by the Member States in these areas. Among the recommendations the following can be highlighted: gathering information about current and planned digitisation and web harvesting of digital resources, both Hispana's objectives.

Hispana consists of four applications: a directory of projects and digitisation initiatives in Spain, an OAI-PMH harvester of digital resources that allows querying the records from these projects and providing access to the digitised document; an OAI-PMH repository and the Search/Retrieve via URL (SRU) server.

The 29th August of 2014 Hispana exceeded 5 million digital objects from 207 OAI-PMH repositories. Hispana has taken part of Europeana since its creation. As of September 2014, there are more than 2 240 000 records from Hispana in Europeana – accounting for 85% of the overall Spanish data contribution. This makes Hispana the second largest Europeana data provider after The European Library.

Due to the large number of libraries contributing to Hispana, 81.8% of the digital objects are textual, and only 17.9% are images. Audio and video represent only 0.3%. Hispana chose to implement the Europeana API in its portal to provide multimedia resources from other cultural institutions alongside its own collections Hispana harvests everything that has a link to a digital object: any object, any audience, any subject and any format.

The INSIDDE repository documented in this deliverable is registered in the Hispania directory (see Figure 3).

Nombre:	Insidde
Entidad responsable:	Treelogic S.L.
Dirección:	Parque tecnológico de Asturias, parcela 30 33428 Llanera (Asturias) Att.: José Antonio Sánchez Tel.: 902 28 63 86 E-mail: joseantonio.sanchez@treelogic.com
Descripción:	En Insidde Repository se registran los resultados obtenidos en el proyecto Insidde (http://www.insidde-fp7.eu/). Insidde es una solución innovadora para la captura, tratamiento de imágenes y digitalización de los detalles desconocidos de pinturas (como bocetos previos ocultos en la obra e identificación de pigmentos) y obras de arte en 3D por medio de la tecnología de terahercios en un enfoque único, con el propósito de ir más allá de nuestro conocimiento actual y la adición de valor a los contenidos culturales.
Fechas:	Creación: 2014
Número de control:	DPD20140000111

Figure 3: INSIDDE repository in Hispana.

3.2 Regional Library of Varna

Regional Library of Varna's collaboration with Europeana started in 2008 when it began enriching Europeana's content with digital objects from Bulgaria. The cooperation proved valuable and in November of 2015 there are around 40 000 Bulgarian items available online through Europeana. This collaboration is an initiative for virtual preservation and presentation of the Bulgarian cultural heritage. Regional Library of Varna acts as aggregators of Bulgarian cultural heritage content to Europeana. However, an official recognised national aggregator and a national established institutional infrastructure for aggregation of the Bulgarian cultural heritage to Europeana are still missing in Bulgaria.

At the moment, the Regional Library of Varna uses ESE format as a communication format with Europeana. They just need a single XML file for every digital object provided.

As a Bulgarian aggregator providing data from Bulgarian cultural organisation usually all metadata is in Bulgarian language. But some of the information such as *Europeana:dataProvider* or *Europeana:Provider* is bilingual (Bulgarian and English).

Regional Library of Varna needs every XML file to follow the structure provided (see Annex A). In order to meet the ESE schema requirements, there are several tags which must be provided like *rights*, *date* and *dataProvider*. A link to every object is also needed.

Other tags are up to data provider. Data providers have to decide which fields they want to provide and which information is important to be presented in Europeana.

4 State of the art of repositories

There is a great variety of software offered on the market for integrated management of repositories and open access archives. Some are free and open license, such as DSpace, E-Prints, Fedora, or Zenity Islandora. On the contrary, others need a business license or payment, such as CONTENTdm Digital Commons, DigiTool, EQUELLA Repository, and Open Repository. The INSIDDE project will focus on an open and free solution.

Most of the repositories in the world use DSpace, E-Prints or Fedora software and they are free and open.

In the next subsections there is an overview of the most common repositories and their characteristics.

4.1 Fedora

The Flexible Extensible Digital Object Repository Architecture (Fedora [10]) is an open source digital repository software, very flexible and capable of serving for a variety of different purposes, such as institutional repositories, content management systems, digital libraries and archives. It is able to store any type of digital content, such as documents, images, video, plus metadata about the content items in any format.

Fedora is a repository maintained and supported by an active community from around the world. It is a product of Fedora Commons, a non-profit organisation, and is provided as free, open-source software. The current development of Fedora is carried out through the non-profit DuraSpace Foundation³, a body set up in 2009 to oversee the sustainable on-going community-based development of two open source digital repository systems, DSpace and Fedora. Fedora project began as a research project at Cornell University's Digital Library Research Group in 1997, which had the aim to research what a system for managing any type of digital content would look like starting from scratch. Fedora development continued on this project basis until 2003, when, together with the University of Virginia, the software was released as a stable production version for a digital repository platform.

It allows storing relationships between content items and also the metadata and relationships for content held by other systems or organisations. Content items can either be stored locally in the repository, or stored externally and just referenced by the Fedora digital object.

Moreover, although it is able to operate as a standalone content server, it is really designed to be used with other software. In most cases it will be only part of a complete content solution incorporating other components such as ingest applications, search engines, workflow management and security.

Fedora key features are the following:

- It can be scaled up easily. The Fedora repository enables the administrators to link to content stored in multiple locations, not only on the repository itself.
- Fedora can manage all kind of content and curate the files regardless of the file format.
- It facilitates interoperability between systems. It makes use of open standards to ensure that the repository can ingest content and can also export it. It uses the standard OAI-PMH to facilitate harvesting.
- It supports the management of related items and allows describing the relation between digital objects. Fedora supports Resource Description Framework (RDF) since it was created. It is able to manage XML-based content as well if preferred.

4.2 DSpace

DSpace [15] is an information system with digital repository architecture focused on the long-term storage, access, preservation and distribution of digital research material in order to ensure the preservation and distribution of all the intellectual production generated within the institutions.

DSpace is an open source platform that is free and can be customised according to the needs of the institution. It is a joint effort between developers from the Massachusetts Institute of Technology (MIT) and Hewlett-Packard Co. The initial goal was to create a scalable and sustainable system capable of hosting more than 100 000 units of digital content produced each year by professors and MIT researchers: articles, reports, communications, also databases, computer programs, video recordings, presentations used in classes, etc.

The software is developed by the MIT, it is open source and freely available to any institution wishing to manage their digital production efficiently. This software was developed using existing rules and standards allowing to easily integrating other information systems. MIT expects to expand its use in other universities with the purpose of creating a "world federation" of digital libraries in the near future.

The keys of the success of Dspace are the following [11]:

- The largest community of users, with more than 1000 implantations.
- As an open source software, there are over 100 developers collaborating with Wiki support and code improvements.
- There is a DSpace Users Group in Spain (GUDE – [12]) where DSpace experts give support to the community and, therefore, and it is very useful to see implantations and keep abreast of software development.
- It can be translated to over 20 different languages, including Spanish and it can also be configured in several languages at once.
- It is used by a wide spectrum of public institutions and private companies. In public administration, many archives, libraries, museums and universities have opted for using this tool.
- Installation and launching is well documented and the parameterisation is not overly complex in a Windows, Linux or Mac OSX environment.
- Supports all types of digital content and an extensive repertoire of formats, both proprietary and free.
- It is an easy and intuitive tool for the end-user.
- It preserves, manages and distributes all types of digital assets.
- It has a system of unique network ID for each item. This ID never change although the URL changes. In order to enable this feature, DSpace uses the Handle system [13].

DSpace can be customised by each organisation and it is worth noting that:

- It has the ability to describe contributions to the repository using the Dublin Core metadata scheme.
- It has the ability to extend the metadata scheme to institutional requirements.
- Users can customise the user interface and offer a differentiated and unique picture of the repository.
- It offers fast, simple and advanced searches against the indexed metadata. Navigation through the collections and resources, and the search engine can also be customisable.
- DSpace meets many of the access and interoperability standards such as OAI-PMH, Open Archives Initiative Object Reuse and Exchange (OAI-ORE), Simple Web-service Offering Repository Deposit (SWORD), OpenSearch, RSS, etc.

4.3 EPrints

EPrints [14] is free and open-source software developed by the University of Southampton. This repository collects, preserves and disseminates the research output created by a research organisation in digital

format. It allows managing the life cycle of the digital object using a web interface, and organises the publications for easy retrieval.

EPrints is a customisable content management system. It can be configured to meet the needs of academics and researchers in reporting and dissemination, but it could be used to manage other digital resources such as images, audio files, etc., definitively, any resource that can be stored digitally.

The key point of this tool is its ease of use for administrators and end-users. The submitting process is straightforward. Users can provide metadata information via web form and manage the lifecycle of the submission in a simple way. They do not need XML or HTML knowledge. On the other hand, administrators can customise the form, so that only the fields selected by them will appear in the form. Administrators also control the browsing categories available and the tool implements multiple browsing criteria including browsing based on the metadata fields.

Moreover, documents in an EPrints repository can be indexed to allow retrieval by search engines like Google. This implies greater access and dissemination of the items stored in the archive. However, one of the weaknesses of EPrints is its inefficient search ability: it is easy to run a search returning no results.

Some features of EPrints are the following:

- Easy to use and administer via web browser.
- Searching based on metadata. It does not support full text based search.
- It uses open source software: Apache and MySQL.
- EPrints meets the standards of interoperability. It offers OAI-PMH support. It allows sites to harvest the metadata from several sources, and offer services using that metadata, such as indexing or linking services.
- Multilingual Support: Unicode is used throughout the software, allowing any language to be processed in a consistent manner.
- It supports file formats such as PDF, HTML, JPEG, TIFF, MP3, AVI, etc.

4.4 Election of repository software

As it was indicated at the beginning of this section, DSpace, Fedora or EPrints support most of the repository installed around the world and they are open license and free.

The software selected for the INSIDDE repository is DSpace because of the following reasons:

- It facilitates the capture and the inclusion of digital objects, including their metadata.
- It facilitates the access to the digital objects in a simple way.
- It facilitates long-term preservation of such digital objects.
- It facilitates the dissemination of the digital objects.
- It implements the OAI-PMH protocol needed to interact with the aggregators.
- Hispana, the National Aggregator in Spain for Europeana, has great expertise in DSpace. They recommended the consortium to use this tool and can offer support in case of any difficulty arises.
- The GUDE can provide support on DSpace.

DSpace serves the purpose of the INSIDDE repository for managing digital objects and meets the purpose of preservation, ingestion, access and dissemination of collections and the publication of the digital objects resulting from the INSIDDE project in Europeana.

5 The INSIDDE repository

5.1 Installation of DSpace

One of the main objectives of task 5.2 “Integration of digital models into Europeana” is to publish the outcomes of the project in the Europeana platform. In order to achieve this goal, the metadata must be mapped to the ESE scheme as specified in previous sections and D5.2.1 [7]. One of the fastest and most effective ways to achieve the objective is to adapt the OAI plugin proposed by Vangelis Banos (documented in the section 5.3) to export the data to the ESE format so that it can be ingested by the aggregators. This plugin is coded in Java and only works with versions prior to the version 3 of DSpace. The consortium was running out of time and was in serious risk of not manage to publish the outcomes of the project in Europeana in month 36. Furthermore, the features and specifications of version 1.8.2 of DSpace were studied and the conclusion was that they met the requirements of the INSIDDE repository. Thus, the version 1.8.2 of DSpace was chosen for the repository.

The source code of DSpace 1.8.2 can be downloaded from [16] via SourceForge [17] and the full documentation is accessible through its wiki [18]. DSpace is a complex application made up by several modules and is managed via web. It supports the most common web browsers such as Mozilla Firefox, Google Chrome, Safari and Internet Explorer.

5.1.1 Software pre-requisites

In order to make DSpace work, the server must have the following software installed. All this software is Open Source:

- *Unix-like OS.* Debian Whezzy v7.8 has been chosen as operating system.
- *Java SDK.* DSpace requires Java 6 or greater due to the incorporation of new language capabilities in 5 and 6 that make coding easier and cleaner.
- *Apache Tomcat.* DSpace requires Tomcat. It is an open-source web server. Tomcat is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation. Version 7 of Tomcat was selected.
- *Apache Maven.* It is a software project management and comprehension tool. Maven is necessary in the first stage of the build process to assemble the installation package for the DSpace instance. Maven v3 was selected.
- *Apache Ant.* Ant is a Java library and command-line tool. Ant's mission is to drive processes described in build files as targets and extension points dependent upon each other. Ant v1.8.2 was selected.
- *PostgreSQL.* PostgreSQL is a powerful, open source object-relational database system. It is required to store the DSpace database. PostgreSQL v9.1.16 was selected.

5.2 Technical specifications

The repository is running in a Debian server and can be accessed through the following URL: <http://idipre.treelogic.com/xmlui/>. Table 1 and Table 2 summarise the technical specifications of the server where the INSIDDE repository is deployed and running.

Table 1: Hardware specifications.

Type	Unit	Quantity
Memory	GB	2
CPU	Core	2

HDD System	GB	15
------------	----	----

Table 2: Software specifications.

Type	Technology	Version
OS	Debian Whezzy	7.8
Repository Software	DSpace	1.8.2
Servlet container	Apache Tomcat	7
Language	Java	1.7.0_79
Build Tool	Apache Maven	3.0.4
Build Tool	Apache Ant	1.8.2
ORDBMS	PostgreSQL	9.1.16

5.3 Customising Dspace

This section details the different processes carried out in order to adapt the repository to the needs of the INSIDDE project. The processes specified consist of the installation of the Handle server for assigning, managing, and resolving persistent identifiers for the digital objects stored in the repository; a tool to generate the thumbnails of the objects and the plugin to export the metadata encoded in Dublin Core to the ESE format.

5.3.1 Persistent identifiers of digital objects

The main reason behind the use of unique and persistent identifiers of the items such as <http://hdl.handle.net/11689/4>, is to provide users who upload objects with a stable identification of such objects. This stable identification can be referenced over time.

To DSpace persistence is interpreted within the meaning of RFC 1737 IETF standard [19], which requires two conditions to Uniform Resource Name (URN):

- Global uniqueness: The same URN will never be assigned to two different resources.
- Persistence: It is intended that the lifetime of a URN be permanent. That is, the URN will be globally unique forever, and may well be used as a reference to a resource well beyond the lifetime of the resource it identifies or of any naming authority involved in the assignment of its name.

DSpace meets these requirements by using two different mechanisms. It can be seen in the following example with <http://hdl.handle.net/11689/4>:

- */4*: this is the external identifier assigned by DSpace. These identifiers are assigned consecutively since the community is created.
- *hdl.handle.net/11689*: the "resolver" that ensures the persistence. Each Dspace instance requires a separate mechanism to perform the location of the persistent identifiers. DSpace uses the CNRI Handle System [20]. Repository administrators have to check in, pay and the Handle System provides an identifier for the institution.

The steps to install a handle server in DSpace are as follows:

1. Perimeter security: Opening 2641 and 8000 TCP/UDP ports to perform the registration process with the entity providing the handle: CNRI.
2. Requesting a handle to handle.net. It is needed to fill in a questionnaire and carry out the user registration (payment) to obtain a handle number to be used in DSpace.
3. Once handle provides the identifier, the handle server is created by means of the following command executed in a terminal:

```
[dspace]/bin/dspace make-handle-config  
[dspace]/handle-server
```

4. Once the server is created, on the folder handle-server there is a .zip file that has to be sent by email to be validated.
5. It is needed to configure DSpace:

- In the file dspace.cfg ([dspace-dir]/dspace/config/dspace.cfg):

```
handle.prefix= 11689
```

- In the file config.dct ([dspace]/handle-server/config.dct):
 - Edit the resulting file to include the following lines in the "server_config" clause:

```
"storage_type" = "CUSTOM"  
"storage_class" = "org.dspace.handle.HandlePlugin"  
Look for "300:0.NA/YOUR_NAMING_AUTHORITY". Replace YOUR_NAMING_AUTHORITY with  
the assigned naming authority prefix sent by CNRI
```

6. Start the handle server :

```
[dspace]/bin/start-handle-server
```

5.3.2 Thumbnails in DSpace

Europeana has developed a policy to improve aspects of the data submitted. They specified the standards for the source images that should be provided for the creation of small images to be used in the portal.

To give users a preview of the objects they have found and to make an attractive interface to the portal, Europeana creates images in two sizes from the objects providers submit, which must be of a reasonable quality. They are generated from the source images whose links are provided in the metadata (in the *europaena:object* tag of the ESE schema). Note that no high quality or large format source objects are stored in the Europeana system.

The requirements of submitting the thumbnail images to Europeana are detailed in D5.1.1 [21].

5.3.2.1 Image Thumbnails

DSpace includes a JPEGFilter as default filter for the tracking of GIF, JPEG and PNG formats.

JPEGFilter filters image bitstreams and scales image to be within the bounds set by the administrator. It creates only JPEGs.

1. In order to enable the filter, the administrator must uncomment the following lines in the settings file of Dspace located in [dspace]/config/dspace.cfg:

```
webui.browse.thumbnail.show  
webui.browse.thumbnail.max.height  
webui.browse.thumbnail.max.width  
webui.item.thumbnail.show  
webui.browse.thumbnail.linkbehaviour
```

```

thumbnail.maxwidth
thumbnail.maxheight
plugin.sequence.org.dspace.app.mediafilter.MediaFilter = org.dspace.app.mediafilter.PDFFilter,
org.dspace.app.mediafilter.HTMLFilter

```

2. Restart the server.
3. Execute the media filters by means of the following command:

```
[dspace]/bin/dspace filter-media
```

4. Create a scheduled task to be executed every day by means of the following command:

```
0 3 * * * [dspace]/bin/filter-media
```

5.3.2.2 PDF Thumbnails

In order to create thumbnails from a PDF files, the software XPDF is required. XPDF is an alternative suite of MediaFilter plugins that offers faster and more reliable text extraction from PDF Bitstreams, as well as thumbnail image generation.

1. Install the XPDF tools [22]:
 - Pdftinfo: displays properties and Info dict.
 - Pdftotext: extracts text from PDF.
 - Pdftoppm: images PDF for thumbnails.
2. Download the Sun Java Advanced Imaging Tools and create a local Maven package:
 - Download the jai_imageio.jar and jai_core.jar files.
 - Install the jar files in the local Maven repository.
3. Edit the DSpace configuration properties to add the location of XPDF executables and reconfigure the MediaFilter plugins. This is needed to set the bounds of the thumbnails.

```

# maximum width and height of generated thumbnails
thumbnail.maxwidth= 160
thumbnail.maxheight = 160

```

4. Add the absolute paths to the XPDF tools installed.

```

xpdf.path.pdftotext = /usr/bin/pdftotext
xpdf.path.pdftoppm = /usr/bin/pdftoppm
xpdf.path.pdftinfo = /usr/bin/pdftinfo

```

5. Change the MediaFilter plugin configuration to remove the old org.dspace.app.mediafilter.PDFFilter and add the new filters.


```

filter.plugins = \
  PDF Text Extractor, \
  PDF Thumbnail, \
  HTML Text Extractor, \
  Word Text Extractor, \
  JPEG Thumbnail
plugin.named.org.dspace.app.mediafilter.FormatFilter = \
  org.dspace.app.mediafilter.XPDF2Text = PDF Text Extractor, \
  org.dspace.app.mediafilter.XPDF2Thumbnail = PDF Thumbnail, \
  org.dspace.app.mediafilter.HTMLFilter = HTML Text Extractor, \
  org.dspace.app.mediafilter.WordFilter = Word Text Extractor, \
  org.dspace.app.mediafilter.JPEGFilter = JPEG Thumbnail, \
  org.dspace.app.mediafilter.BandedPreviewJPEGFilter = Banded Preview JPEG

```

6. Add the input format configuration properties for each of the new filters.

```

filter.org.dspace.app.mediafilter.XPDF2Thumbnail.inputFormats = Adobe PDF
filter.org.dspace.app.mediafilter.XPDF2Text.inputFormats = Adobe PDF
filter.org.dspace.app.mediafilter.HTMLFilter.inputFormats = HTML, Text
filter.org.dspace.app.mediafilter.WordFilter.inputFormats = Microsoft Word
filter.org.dspace.app.mediafilter.PowerPointFilter.inputFormats = Microsoft Powerpoint,
Microsoft Powerpoint XML
filter.org.dspace.app.mediafilter.JPEGFilter.inputFormats = BMP, GIF, JPEG, image/png
filter.org.dspace.app.mediafilter.BandedPreviewJPEGFilter.inputFormats = BMP, GIF, JPEG,
image/png

```

7. Add the filter name in the filter.plugins property.

```

filter.plugins = \
  PDF Text Extractor, \
  PDF Thumbnail, \
  HTML Text Extractor, \
  Word Text Extractor, \
  PowerPoint Text Extractor, \
  JPEG Thumbnail, \
  Banded Preview JPEG

```

5.3.3 Plugin Crosswalk Dublin Core to ESE

The metadata schema selected to store the metadata generated in the project in the INSIDDE repository is Dublin Core. DSpace supports exporting Dublin Core metadata through the OAI-PMH protocol by default.

There is an incompatibility issue because Europeana and, by extension, the aggregators, do not support Dublin Core but ESE – an application profile which is based on Dublin Core – and EDM. The repository must be interoperable with the aggregators; therefore it is necessary to "talk" in the same language. As it is specified in D5.2.1 [7], the selected schema to contribute to Europeana is ESE.

In DSpace, the export to ESE is not supported. Consequently, it is essential to modify DSpace in order to export ESE metadata. This goal is achieved through the inclusion of an OAI plugin based on the one proposed by Vangelys Banos [23], which is used to export ESE-v3.4 metadata. This plugin is coded in Java and it basically implements the mapping from Dublin Core to ESE documented in D5.2.1 [7].

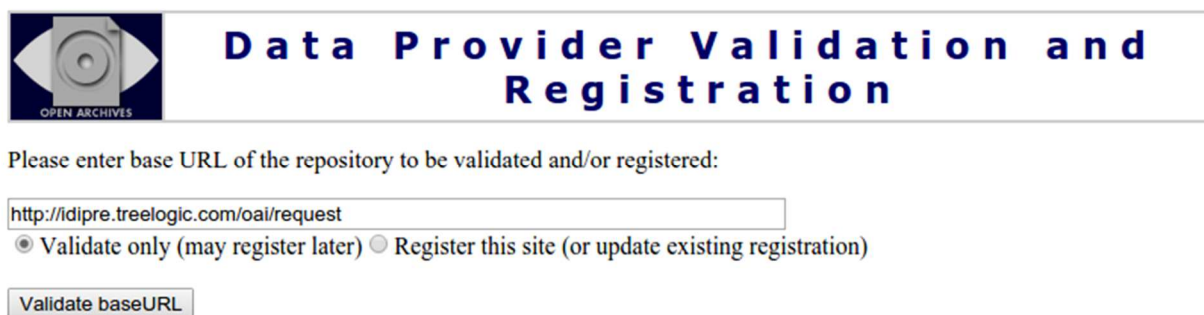
The procedure to install the plugin in DSpace consists of two steps:

1. Compile java-oai.jar with ESE support.
 - Move the java file (ESECrosswalk.java) to the folder: dspace-1.8.2-src-release/dspace-oai/dspace-oai-api/src/main/java/org/dspace/app/oai.
 - Go to the folder: dspace-1.8.2-src-release/dspace-oai/dspace-oai-api and run mvn package to compile the software and create the dspace-oai-api-1.8.2.jar file.
2. Configure Dspace installation in order to enable ESE support in the new java-oai.jar.
 - Copy the file dspace-oai-api-1.8.2.jar in /dspace/webapps/oai/WEB-INF/lib/dspace-oai-api-1.8.2.jar.
 - Modify Dspace configuration file /dspace/config/oaicat.properties and append the following text entry Crosswalks.esse=org.dspace.app.oai.ESECrosswalk.
 - Restart the server.

The content of the URL <http://idipre.treelogic.com/oai/request?verb=ListRecords&metadataPrefix=ese> was revised and validated together with Hispana and it was verified that the INSIDDE repository supports ESE.

5.4 OAI validation

In order to make sure that the repository correctly implements the OAI-PMH, the consortium followed the Open Archive Initiative Protocol Testing.



Data Provider Validation and Registration

Please enter base URL of the repository to be validated and/or registered:

Validate only (may register later) Register this site (or update existing registration)

Figure 4: OAI-PMH validation.

The Protocol Testing of OAI-PMH defines a set of requests and the format of the responses to each of these requests, specified through an XML schema. A conformant repository must respond to each of the protocol requests with a response that validates with the schema. It also must respond to mal-formed requests with the appropriate errors and exception conditions.

The server was validated using the validator supplied by the OAI [24] (see Figure 4). The validation was successful and the repository has passed the initial validation as can be seen in Figure 5.



Data Provider Validation and Registration

Initial validation checks

baseURL is <http://idipre.treelogic.com/oai/request>
Validation only
Request logged from 85.152.52.194

Checking Identify response

Request: GET <http://idipre.treelogic.com/oai/request?verb=Identify>
Administrator email address joseantonio.sanchez@treelogic.com
[PASS] Correctly reports OAI-PMH protocol version 2.0
[PASS] baseURL supplied matches the Identify response
[PASS] Datestamp granularity is 'seconds'
[PASS] earliestDatestamp is 2001-01-01T00:00:00Z

This repository has passed initial validation. An email message has been sent to the administrator's email address shown above which includes a URL to continue validation and/or registration. The purpose of this is to ensure that only the repository administrator can change the registration record of a repository, and to avoid the excessive use of the validation script by some third party.

Figure 5: OAI-PMH validation result.

6 Content and Metadata Harvesting

The INSIDDE repository [25] consists of two collections: one housing the Bulgarian objects and the other housing the Spanish objects.

The collection housing the Bulgarian objects contains the 3D models (see Figure 6) of four vessels property of RIM Stara Zagora belonging to the Roman Empire (I-III A.D). The 3D models are in 3D-PDF format as detailed in D5.1.2 [26]. The templates containing the metadata (see D5.2.1 [7]) were sent by RIM Stara Zagora on 28 September 2015 and were immediately added to the repository. The metadata is written in Bulgarian language as it was required by the Regional Library of Varna, the Bulgarian aggregator.

The template sent by the Bulgarian aggregator was filled with the metadata in ESE format and submitted on 8 October 2015. According to the Regional Library of Varna, the content should be published in Europeana by the end of 2015, just before the end of the project.

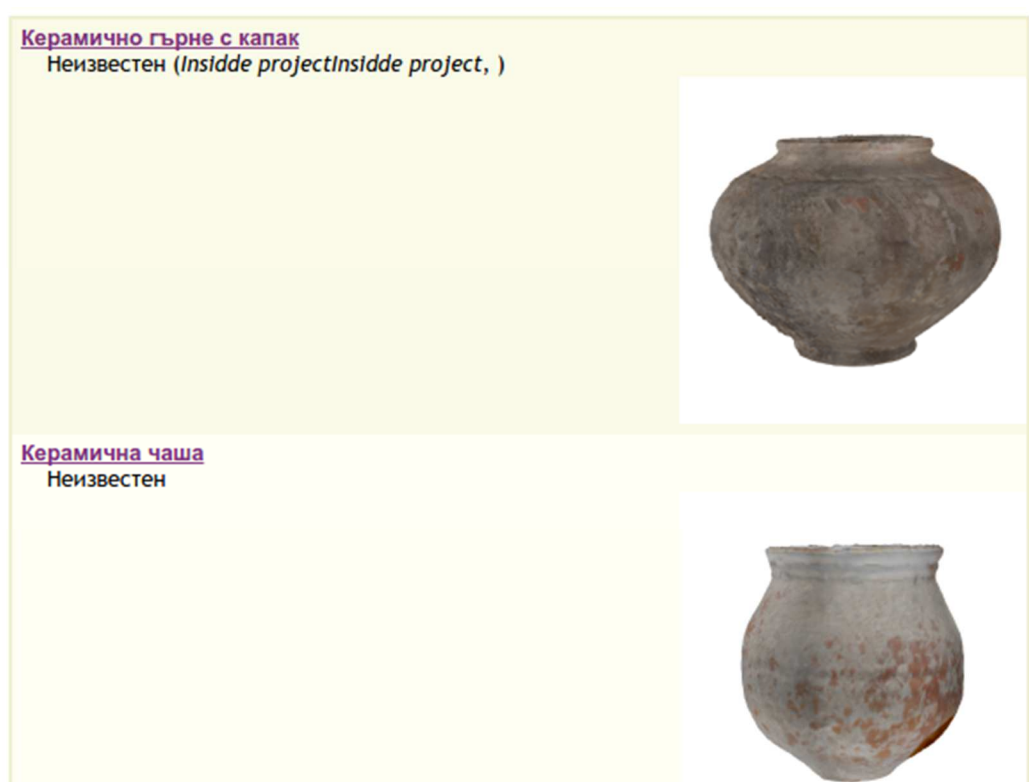


Figure 6: INSIDDE repository (Bulgarian collection).

The collection housing the Spanish objects (see Figure 7) contains 2D images of 15 paintings belonging to the Museum of Fine Arts of Asturias plus one image resulting from an infrared analysis carried out by CNR-INO. The templates containing the metadata were sent by the Asturian museum at the beginning of October.

Hispana, the aggregator in charge of the Spanish objects, and Treelogic scheduled the ingestion for the night of the 14 October 2015. According to the aggregator, the content should be published in Europeana at the beginning of December 2015.

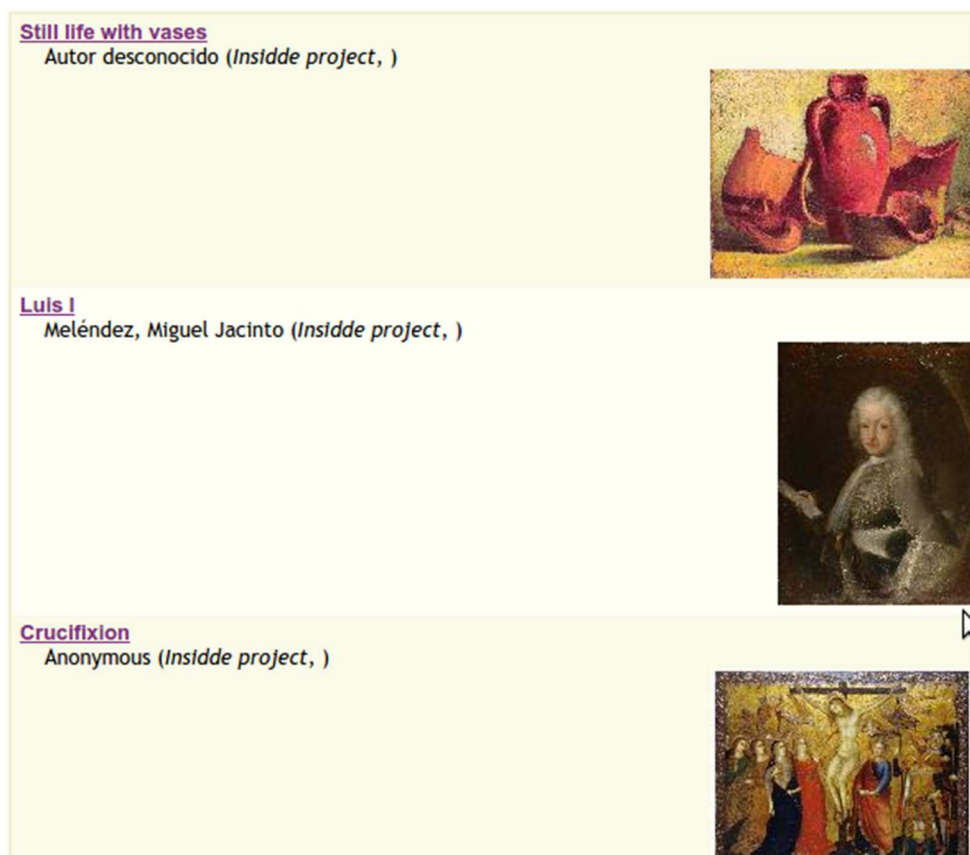


Figure 7: INSIDDE repository (Spanish collection).

7 Conclusions

This report documents the guidelines to follow for integrating the 2D and 3D models resulting from the project into Europeana. An essential requirement to contribute to the cultural heritage platform is to create and set up a repository that provides a mechanism to manage the digital surrogates and the metadata associated.

The INSIDDE project employs cultural objects from Bulgaria and Spain, therefore, the consortium had to collaborate with the aggregators from both countries that are the channel through which the metadata can be harvested and published on Europeana: Regional Library of Varna, the Bulgarian aggregator, and Hispana, the Spanish National Aggregator.

The Regional Library of Varna need a single XML file in ESE format for every digital object provided but, generally, aggregators harvest the metadata records from repositories by means of the OAI-PMH protocol. This makes necessary to select a repository that implements such protocol.

This report documents the study of the three main free and open license repository platforms in the market: Fedora, DSpace and Eprints.

The selected repository software is DSpace since it fulfils the requirements established, it was recommended by Hispana, and in addition has the GUDE, a Spanish DSpace users group, whose objectives are to share experiences and knowledge about the implementation and support on DSpace.

The deliverable also details the adaptation of the repository to meet the goal of integrating the digital objects into Europeana. The adaptation mainly consists of the installation of the Handle server for assigning, managing, and resolving persistent identifiers for the digital objects stored in the repository, development of a tool to generate the thumbnails of the objects, and finally the implementation of the plugin to export the metadata encoded in Dublin Core to the ESE format that was specified in D5.2.1 [7].

The INSIDDE repository, which can be accessed through <http://idipre.treelogic.com/xmlui/>, consists of two collections: one housing the Bulgarian objects and the other housing the Spanish objects.

The Bulgarian collection contains the 3D models of four vessels property of RIM Stara Zagora. In turn, the Spanish collection contains 2D images of 15 paintings belonging to the Museum of Fine Arts of Asturias plus one image resulting from an infrared analysis carried out by CNR-INO. The metadata was harvested by both aggregators in the first half of October. According to the aggregators, the content is planned to be published in Europeana in December 2015.

References

- [1] <http://www.budapestopenaccessinitiative.org/read>
- [2] <https://www.openarchives.org/pmh/>
- [3] Semple, N. (2006). Digital repositories. Retrieved from <http://www.dcc.ac.uk/resources/briefingpapers/introduction-curation/digital-repositories>
- [4] <https://www.openarchives.org/>
- [5] Lynch, Clifford A. (January 2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." ARL Bimonthly Report 226. Available: <http://www.arl.org/newsltr/226/ir.html>
- [6] Armbruster, C., & Romary, L. (2010). Comparing Repository Types - Challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication. Digital Libraries. Retrieved from <http://arxiv.org/abs/1005.0839>
- [7] *D5.2.1 Report on compilation of metadata for 2D/3D artworks in Europeana (ongoing)*. Jose Antonio Sánchez and others. INSIDDE
- [8] Europeana Professional portal. Online (1 December 2015): <http://pro.europeana.eu/>
- [9] Hispana. Online (1 December 2015): <http://hispana.mcu.es>
- [10] <http://www.fedora-commons.org/>
- [11] Garrido, J. (2012). La viabilidad de la implantación de EDM (Europeana Data Model) en los repositorios DSpace (objeto application/pdf). Retrieved August 22, 2013, from http://www.mcu.es/principal/docs/novedades/2012/Europeana6_Javier_Garrido.pdf
- [12] <https://www.rediris.es/list/info/gude.html>
- [13] CNRI. (2013). Handle System. Retrieved September 10, 2013, from <http://www.handle.net/>
- [14] <http://www.eprints.org/>
- [15] <http://www.dspace.org/>
- [16] <http://sourceforge.net/projects/dspace/files/DSpace%20Stable/1.8.2/>
- [17] <http://sourceforge.net>
- [18] <https://wiki.duraspace.org/display/DSDOC18/DSpace+1.8+Documentation>
- [19] <https://www.ietf.org/rfc/rfc1737.txt>
- [20] <https://www.handle.net/>
- [21] *D5.1.1 Report on image processing applied to paintings for end-user applications* (September 2015). Jose Antonio Sánchez, Marcos Fernandez and others. INSIDDE
- [22] <http://www.foolabs.com/xpdf/>
- [23] Banos, V. (2010). *Dspace plugin for Europeana Semantic Elements*. <http://vbanos.gr/blog/2010/02/11/dspace-plugin-for-europeana-semanticelements/>
- [24] <http://www.openarchives.org/data/registerasprovider.html>
- [25] <http://idipre.treelogic.com/xmlui>
- [26] *D5.1.2 Demonstration on 3D artwork models using the created end-user applications* (September 2015). Jose Antonio Sánchez and Marcos Fernandez. INSIDDE

Annex A Regional Library of Varna ESE template

```

<?xml version="1.0" encoding="UTF-8"?>
<europeana:metadata
  xmlns:europeana="http://www.europeana.eu/schemas/ese/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.europeana.eu/schemas/ese/
http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd">
<!-- Europeana tags -->
<europeana:record>
<!-- Strongly recommended elements - SR-->
<dc:title>Eight weeks</dc:title>
<dcterms:alternative>Ocho seman</dcterms:alternative>
<dc:creator>Shakespeare, William</dc:creator>
<dc:contributor>Maria Callas</dc:contributor>
<dc:date>1564</dc:date>
<dcterms:created>1883-12-24</dcterms:created>
<dcterms:created>Iron Age</dcterms:created>
<dcterms:issued>1993</dcterms:issued>

<!-- Recommended elements - R -->

<dc:coverage>1995-1996</dc:coverage>
<dc:coverage>Boston, MA</dc:coverage>
<dcterms:spatial>Portugal</dcterms:spatial>
<dcterms:temporal>Roman</dcterms:temporal>
<dc:description>Illustrated guide to ...</dc:description>

<dcterms:isPartOf>Crace Collection of Maps of London</dcterms:isPartOf>
<dc:language>bg</dc:language>
<dc:publisher>Oxford University Press</dc:publisher>
<dc:source>Security Magazine pp 3-12</dc:source>
<dc:subject>submarine</dc:subject>
<dc:type>photograph</dc:type>

<!-- Some of the additional elements -->

<dc:format>image/jpeg</dc:format>
<dcterms:extent>13 cm</dcterms:extent>
<dc:identifier>
http://upload.wikimedia.org/wikipedia/en/f/f3/Europeana_logo.png</dc:ide
ntifier>
<dc:rights>Creative Commons Attribution 3.0 License</dc:rights>
<dcterms:provenance>Donated by The National Library in
1965</dcterms:provenance>
<dc:relation> maps.crace.1/33</dc:relation>

```

```
<!--Europeana Elements OBLIGATORY -->

<europeana:object></europeana:object>
<europeana:provider>Регионална библиотека ПЕНЧО СЛАВЕЙКОВ - Варна /
Public Library - Varna</europeana:provider>
<europeana:type>3D</europeana:type>
<europeana:rights>http://creativecommons.org/publicdomain/zero/1.0/</eur
opeana:rights> <!-- according to CC -->
<europeana:dataProvider>Регионален исторически музей - Стара Загора /
Regional Museum of History - Stara Zagora</europeana:dataProvider>
<europeana:isShownBy></europeana:isShownBy>
<europeana:isShownAt></europeana:isShownAt>

</europeana:record>
</europeana:metadata>
```