



Europeana Space – Spaces of possibility for the creative reuse of Europeana’s content  
CIP Best practice network - project number 621037

<b>Deliverable number</b>	D2.1
<b>Title</b>	Requirements for the creative use of Digital Cultural Resources; progress on collaboration towards Europeana Labs

<b>Due date</b>	Month 9
<b>Actual date of delivery to EC</b>	1 December 2014

<b>Included (indicate as appropriate)</b>	Executive Summary	<input checked="" type="checkbox"/>	Abstract	<input type="checkbox"/>	Table of Contents	<input checked="" type="checkbox"/>
---	-------------------	-------------------------------------	----------	--------------------------	-------------------	-------------------------------------

The change in title of the deliverable and extension to the due date, were agreed with the Project Officer.

**Project Coordinator:**

Coventry University

Prof. Sarah Whatley

Priority Street, Coventry CV1 5FB, UK

+44 (0) 797 4984304

E-mail: [S.Whatley@coventry.ac.uk](mailto:S.Whatley@coventry.ac.uk)

Project WEB site address: <http://www.europeana-space.eu>

### Context:

<b>Partner responsible for deliverable</b>	NTUA
<b>Deliverable author(s)</b>	Nasos Drosopoulos
<b>Deliverable version number</b>	1.0

<b>Dissemination Level</b>	
<b>Public</b>	<input checked="" type="checkbox"/>
<b>Restricted to other programme participants (including the Commission Services)</b>	<input type="checkbox"/>
<b>Restricted to a group specified by the consortium (including the Commission Services)</b>	<input type="checkbox"/>
<b>Confidential, only for members of the consortium (including the Commission Services)</b>	<input type="checkbox"/>

### History:

<b>Change log</b>			
<b>Version</b>	<b>Date</b>	<b>Author</b>	<b>Reason for change</b>
0.1	02/05/2014	Nasos Drosopoulos, Antonella Fresa, Stefanos Kollias	Original D2.1 "Requirements for the creative use of Europeana Cultural Resources" modified to also include "Report on joint development of Europeana Lab with Europeana Creative"
0.2	22/07/2014	Nasos Drosopoulos, Tim Hammerton, Sarah Whatley, Antonella Fresa Vassilis Tzouvaras	Deliverable rescheduled for M9. New title, to incorporate Task 1.5 input, "Requirements for the creative use of Digital Cultural Resources; progress on collaboration towards Europeana Labs"

0.5	30/08/2014	Nasos Drosopoulos, Natasa Sofou	Technical questionnaire for pilots
0.7	9/10/2014	Nasos Drosopoulos, Vassilis Tzouvaras	First draft
0.8	16/10/2014	Nasos Drosopoulos, Vassilis Tzouvaras	Outline presented and discussed in Venice Plenary
0.9	24/11/2014	Nasos Drosopoulos, Tim Hammerton, Antonela Fresa, Nikos Simou, Giorgos Marinellis, Alexandros Chortaras	Peer review version T1.5 input included Update of requirements and architecture
1.0	01/12/2014	Nasos Drosopoulos	Final version

<b>Release approval</b>			
<b>Version</b>	<b>Date</b>	<b>Name &amp; organisation</b>	<b>Role</b>
1.0	1/12/2014	Tim Hammerton, COVUNI	Project Manager

**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.

## TABLE OF CONTENTS

<b>1</b>	<b>EXECUTIVE SUMMARY</b> .....	<b>7</b>
<b>2</b>	<b>INTRODUCTION</b> .....	<b>8</b>
2.1	BACKGROUND.....	8
2.2	ROLE OF THIS DELIVERABLE IN THE PROJECT.....	9
2.3	APPROACH .....	9
2.4	STRUCTURE OF THE DOCUMENT .....	10
<b>3</b>	<b>DIGITAL CULTURAL RESOURCES IN EUROPEANA SPACE PILOTS</b> .....	<b>11</b>
3.1	QUESTIONNAIRE RESULTS .....	11
3.1.1	<i>Content</i> .....	11
3.1.2	<i>Metadata</i> .....	12
3.1.3	<i>Services</i> .....	13
3.1.4	<i>Operational scenarios and planning</i> .....	13
<b>4</b>	<b>TECHNICAL SPACE REQUIREMENTS</b> .....	<b>15</b>
4.1	FUNCTIONAL REQUIREMENTS.....	15
4.1.1	<i>Access to content</i> .....	15
4.1.2	<i>Access to metadata</i> .....	15
4.1.3	<i>Dataset validation, transformation &amp; versioning</i> .....	15
4.1.4	<i>Dataset Indexing and statistics</i> .....	16
4.1.5	<i>Aggregate multiple sources</i> .....	16
4.1.6	<i>Identification, authentication and IPR control</i> .....	16
4.1.7	<i>Metadata cleaning and enrichment</i> .....	17
4.1.8	<i>Semantic web serializations &amp; technologies</i> .....	17
4.1.9	<i>Multi-purpose access APIs</i> .....	17
4.2	NON-FUNCTIONAL REQUIREMENTS .....	18
<b>5</b>	<b>TECHNICAL SPACE ARCHITECTURE</b> .....	<b>19</b>
5.1	STORAGE .....	20
5.1.1	<i>Metadata</i> .....	20
5.1.2	<i>Content</i> .....	21
5.1.3	<i>Importing and Harvesting</i> .....	21
5.2	DIGITAL CULTURAL RESOURCES AGGREGATION AND MANAGEMENT .....	21
5.2.1	<i>Aggregation and Publication workflow</i> .....	21
5.2.2	<i>Processing Infrastructure</i> .....	22
5.2.3	<i>Service Integration and Deployment</i> .....	22
5.3	CONTENT REUSE FRAMEWORK .....	22
5.4	METADATA PROCESSING UNIT .....	23
5.5	ACCESS APIs .....	23
<b>6</b>	<b>PROGRESS ON COLLABORATION TOWARDS EUROPEANA LABS</b> .....	<b>25</b>
6.1	BACKGROUND.....	25
6.2	RELATIONSHIP.....	26
6.3	SOURCES OF DIGITAL CONTENT .....	27
6.4	NEXT STEPS.....	28
<b>7</b>	<b>PLANNING</b> .....	<b>29</b>

<b>8</b>	<b>SUPPORT .....</b>	<b>30</b>
<b>9</b>	<b>CONCLUSION .....</b>	<b>31</b>

## 1 EXECUTIVE SUMMARY

The Europeana Space project aims to increase and enhance the creative industries' use of Europeana and other online collections of digital cultural content, by delivering a range of resources to support their engagement. The project addresses all sectors of the creative industries, from content providers to producers, exhibitors, artists and makers of cultural/creative content, publishers, broadcasters, telecoms and distributors of digital content.

The Technical Space will be available for cultural institutions and organisations, professional users and third party developers in order to easily search for the cultural resources that meet their retrieval criteria so as to use and re-use them for the development of applications. This will be achieved through the delivery of APIs that will facilitate the development of applications based on cultural content and the realisation of the six Pilot projects. Additional requirements will be produced in the Innovation Space, such as the six Pilot related hackathons, and will be evaluated and addressed by the Technical Space, reflecting a 'real-world' approach to development that can be made immediately useful.

This document reports on the collection and analysis of requirements with respect to content and metadata, formats, types of access and operational scenarios. The aim is to further specify the functionalities of the APIs developed and to outline the architecture of the Technical Space. It is also crucial to report and show the established and potential cooperation with other activities in the Europeana ecosystem in order to ensure the interoperability of available platforms and the complete representation of all stakeholders' requirements and expectations.

Finally, the complementary strategy of the Europeana Foundation is considered in their reaching out to the creative industries and how one of the key planks in this is the development of Europeana Labs. Options of how Europeana Space may interact with the Europeana Foundation and the Europeana Creative project, towards cooperation in the coordination and potential joint development of the Europeana Labs are considered, as well as opportunities for the Europeana Space project to utilize existing datasets that are already available within Labs to provide content for project activity.

As with all project activity, the challenge of finding suitable digital cultural resources will vary from pilot to pilot in terms of available content and the requirements of project stakeholders. This work has already begun within the Technical Space to support project activity in other work packages within the wider Europeana ecosystem context.

## 2 INTRODUCTION

### 2.1 BACKGROUND

Cultural Heritage has been the focus of a great and continually increasing number of research and development projects and initiatives during the last decade, aiming at efficiently managing and disseminating cultural resources on the Web. As more institutions make their digitized collections available online, and proceed to aggregate them in domain repositories, they reinforce the growing number of cultural heritage sources and resources available to users. In the last couple of years, interaction with the industry combined with efforts towards IPR definition, clearance and rights labeling have enabled the distribution of content available for re-use. Europeana in particular is contributing to the open data realm by releasing aggregated metadata under a CC0 Public Domain Dedication. A growing network of cultural heritage professionals and scholars, software designers and developers, interaction and data experts, promote and stimulate the creation of innovative applications based on available data.

Users of cultural heritage repositories are primarily directed to navigation and presentation interfaces where they can search, browse and visualize the results of digitization and aggregation activities. As with any environment that involves a significant number of distributed sources that produce or publish content at an increasing rate, quality of user experience over time may deteriorate. The interfaces often lack the ability to intuitively present large result sets and, as is the case with digital cultural heritage material, the complexity and detail of available data structures, such as collections and hierarchies of material, are neither indexed nor rendered efficiently. The content itself, along with the various publication strategies of providers and aggregators also influence user experience, as one can observe great differences in metadata expressivity and quality.

In the case of Europeana, there is content with basic metadata, other with expressive structures that are merged or simulated by using, or misusing, the Europeana Data Model, and in general, there is great variety in both knowledge quality and its resulting visualization within the results of a single user search. As aggregators evolve by introducing and focusing on detailed, domain specific curation, presentation and publication requirements, it becomes harder to coherently navigate between different contributions, an experience already inhibited by variations in intrinsic features such as multi-linguality and copyright approaches.

It is then those same features that denote the significance and richness of cultural heritage, such as diversity, growth and dispersion of sources, which maintain, or raise a wall that inhibit its transition and transformation. Re-use of the Europeana content and promotion of creativity and, more importantly, co-creation are among the main targets for the community, as it attempts to open up to the industry and include more citizens in the development and appreciation of digitized cultural heritage. Even that goal itself introduces obstacles, as participants of co-creation are identified as a broad array of diverse users, such as recreational, education and business.

In this environment, knowledge-based management and retrieval becomes a necessary evolution from simple syntactic data exchange. In the process of aggregating heterogeneous resources and publishing them for retrieval and creative re-use, networks such as Europeana and DPLA invest in technologies that achieve semantic data integration. The resulting repositories join the Linked Open Data cloud, allowing to link cultural heritage domain knowledge to existing datasets. Integration of diverse information is achieved through the use of formal ontologies, enabling reasoning services to offer powerful semantic search and navigation mechanisms. This, in turn, can empower the efforts towards re-using the content in a wide range of applications for the creative industry.



Europeana Space aims at the establishment of the Technical Space as a framework for storing, accessing and processing content and metadata. It will be interoperable with complementary services emerging from the Europeana group of projects and will take advantage of respective infrastructures being established, such as the Europeana Cloud. Cultural institutions and organisations, professional users and third party developers will be able to easily search for cultural resources meeting their retrieval criteria so as to use and re-use them for the development of applications.

Its data infrastructure consists of:

- a content retrieval system to complement content sourcing for applications,
- a core repository and services for metadata ingested to the platform and,
- a semantic repository for the respective serialisation of those resources and their combination with ones extracted from other sources (primarily Europeana but also other DCH repositories like DPLA, or crowd-sourced repositories such as audiovisual channels).

The Technical Space will include appropriate APIs to facilitate the development of applications based on cultural content starting from the realisation of the six Pilot projects. It will also enable the integration of services for further processing of content or metadata, starting from the metadata processing unit that will offer aggregation services for the alignment and enrichment of DCH resources and, continuing with services developed within the Pilots or provided via Europeana Labs.

## **2.2 ROLE OF THIS DELIVERABLE IN THE PROJECT**

This document reports on the steps taken for the collection of requirements that guide the development of the Technical Space, (in alignment of the activities of WP3 Content Space, WP4 Pilot and WP5 Innovation Space). It discusses functional and non-functional requirements gathered internally, from the coordination and development teams for the project's 6 pilots, as well as from a variety of relevant projects and activities within the Europeana Network.

This leads to a more detailed specification of the Technical Space's architecture while also identifying potential synergies and cooperation with existing and under development systems, in various technical levels such as storage infrastructure or third party services for developers and users. Interoperability with the Europeana ecosystem is a significant aspect of the architecture in order to inform technical choices with existing system capabilities and requirements.

The deliverable also reports on non-technical aspects of collaboration with the Europeana Network, in the form of sustainability approaches and communication initiatives.

Finally, it discusses planning and implementation priorities to achieve short and long term goals, as well as the processes and tools to support developers and users of the Technical Space.

## **2.3 APPROACH**

The results presented in this report are the outcome of a set of activities that involve the Project Management work package, specifically Task 1.5 "Joint development of Europeana Lab infrastructure", WP2 that is responsible for the infrastructure and tools for content access, use

and storage and, the Europeana Space Scenarios WP4, which develops the pilots in the six thematic areas.

In particular, deliverables D4.1 and D4.2 regarding pilot methodology, content sourcing and coordination were crucial for the identification and validation of requirements. For the production of D4.2, WP2 issued a questionnaire during the pilots' planning phase, in order to collect and analyse requirements in terms of:

- Content (type, formats, sources, licensing)
- Metadata (models, serializations, access, transformations required, sources)
- Additional services (pilot development, Europeana Labs, other projects)
- Operational scenarios (web services vs downloadable apps, ongoing or 'one-off' ingestion, remote calls to services, personalization, expected load on the APIs etc.)
- Planning

This analysis and the related meetings and collaborations with pilots during this period finalized the high-level functionalities that will have to be provided through the WP2 APIs so that the Pilots will act as a reference implementation for the use of the infrastructure.

From that point on, constant communication and cooperation with the pilots' teams keep driving implementation, helping to identify evolving requirements as the project also prepares the Innovation Space (WP5) to address the needs of a wider user base and market. The Technical Space must be extendable towards future requirements and services, as the project will move from sandboxing and prototyping towards enterprise development and incubation.

WP2 leader and other project's partners participate in further activities in the Europeana ecosystem, especially in related projects such as Europeana Creative, LoCloud, Europeana Food & Drink and Europeana Sounds. Participation in Europeana's Projects Group Assembly and related technical workshops as well as in the EuropeanaTech R&D community and its respective task forces has also assisted in gathering requirements from researchers, developers and experts in the network.

## **2.4 STRUCTURE OF THE DOCUMENT**

Chapter 3 presents the result of the questionnaires submitted by pilot coordination and development teams regarding the expected usage and desired features of the technical infrastructure. Chapter 4 discusses the functional and non-functional requirements that were identified in Europeana Space, as well as through input collected by various projects and activities within the Europeana Network, while Chapter 5 describes the architecture of the Technical Space. Chapter 6 presents the initiatives and progress on collaboration towards Europeana Labs. Chapter 7 highlights important development and deployment planning information while Chapter 8 presents the tools and processes that are put in place in order to support the development of Pilots initially, and then the activities of the Innovation Space.

### **3 DIGITAL CULTURAL RESOURCES IN EUROPEANA SPACE PILOTS**

In order to explore different scenarios of re-use of Europeana and other available digital cultural content, the Europeana Space project is developing six Pilots in the thematic areas of Europeana TV, Photography, Dance, Games, Open and Hybrid Publishing and, Museums. The pilots (WP4), despite their different focus and objectives, have a common background developed within the project, in particular in the areas of Technical integration (WP2) and of IPR (WP3). Their role in the development and evolution of the Technical Space is dual, as they act both as its first users and at the same time as the driving force for its implementation. Although their development plans are independent of WP2 scheduling and will only require full availability of the Technical Space after the prototyping phase, they act as representatives of the creative industry that the project, and its technical infrastructure specifically, is aiming to cater for. In that sense, they are instrumental in the process of defining the functional and non-functional requirements that will characterize the Technical Space's architecture.

Towards that direction, the Pilot coordinators were invited by WP2 to complete a questionnaire in order to identify requirements for the use of content and metadata, specify the technical aspects of development, integration and reuse of services and, provide operational and planning information. WP2 also participated in several meetings with the pilot teams regarding technical aspects, planning and brainstorming for developing new use-cases for the pilots. The questionnaire was issued during the planning phase for pilots and, while it is representative of the expected usage and methodology, it is expected that during the next phases of pilot development as well as during the Innovation Space events, such as hackathons and incubation, some of the aspects may be expanded or redefined. In conjunction with the requirements identified from the interaction with activities in the Europeana Network, the Technical Space will be able to accommodate future, yet unreported scenarios.

Results of the questionnaire analysis indicate that ease of access and manipulation of metadata resources from distributed repositories is crucial for the establishment of a use case scenario. Aggregation using specific data models and the ability to access different serializations according to expected usage are needed to facilitate design and development approaches. Content, available in medium or higher quality is also essential for many of the applications envisioned. Operational scenarios indicate the need for constant access to data and for authenticated delivery. Fulfilling these requirements, in a platform that facilitates interoperability enables developers to use available advanced services for the manipulation of content and metadata, as well as to contribute and integrate their developed innovative services.

#### **3.1 QUESTIONNAIRE RESULTS**

##### **3.1.1 Content**

The questionnaire addressed the use of cultural heritage content by the pilots, specifically the types and respective formats of digital resources, the identified sources and repositories and the licensing schemes they are expected to support.

The pilots are expected to use four of the 5 available content types reported in Europeana, specifically (number of pilots that responded positively follows each type), Images (6), Video (4), Text (3) and Audio (3), while no pilot expects to use 3D resources. Formats include the prevalent ones available by institutions for the web such as JPEG and PNG for images, video

encodings for streaming such as flash and mp4, PDF and plain text for textual resources and MP3, WAV or AIFF for audio. As most applications are based on the content itself we consider qualities from average and above for images and video.

Sources of content include Europeana (all pilots), content curated by the pilots (from the regional, national and private archival collections of partners) and ingested in European Space (4 pilots) and content from the Digital Public Library of America (2 pilots). Pilots will also investigate the use of content available on public repositories such as

- Digital Commons (<http://digitalcommons.bepress.com/>)
- OA art research archive/repository (<http://www.researchcatalogue.net>),
- National Digital Archive Poland, (<http://audiovis.nac.gov.pl/>),
- Ada – Archive of Digital Art (<https://www.digitalartarchive.at>),
- Sound and vision, NL, Open Images (<http://www.openimages.eu/>),
- Europeana Early Photography (<http://www.earlyphotography.eu/>),
- Wikimedia Commons (a media file repository making available public domain and freely-licensed educational media content such as images, sound and video clips),
- Critical Commons (a public media archive and fair use advocacy network that supports the transformative reuse of media in scholarly and creative contexts),
- academic journals published as part of the Public Knowledge Project's Open Journal System jointly owned by Cov Uni and Siobhan Davies Dance
- the EUScreen project.

Pilots also take advantage of widely used content delivery services such as Vimeo and YouTube for videos and, Flickr for images.

Finally, with respect to licensing the following list shows the detailed replies from the six pilots, highlighting the expected support for multiple licensed content (3 out of 6 pilots):

- Free usage, Require attribution
- Free usage (CC-BY or equivalent)
- Free usage, Non-commercial usage for research/academic, Non-commercial usage
- Free usage, Require attribution
- Free usage, Non-commercial usage for research/academic, Non-commercial usage, Commercial usage, Require attribution
- Non-commercial usage for research/academic, Commercial usage

### 3.1.2 Metadata

The questionnaire addressed the use of metadata describing content, specifically the data models and their respective serializations, the types of access required, expected transformations to other data models and, linking to external terminologies (e.g. Getty vocabularies) and repositories of knowledge (e.g. DBpedia, Geonames).

The results indicate that all pilots will start with metadata following the Europeana Data Model and its available profiles such as the DM2E (<http://dm2e.eu>) for manuscripts and the profile developed in the Europeana Sounds project. As the applications become available and generate user feedback and additional requirements, two of the pilots indicated that they would explore the need for creating a specific profile for Europeana Space.

Required serializations for metadata records include XML, RDF/XML, N-triples and JSON-LD and there are three access scenarios, via HTTP download, HTTP APIs and, SPARQL endpoints.

Sources for enrichment include only the indicated ones (Getty, DBPedia, Freebase and Geonames) as pilots are initially sourcing content with metadata that are already rich in information. Regarding the potential enrichment approach, two of the pilots indicated that only URL links to other sources will be enough, while three would require to retrieve resources from third party repositories in order to include them in the records used by the application.

### 3.1.3 Services

This part of the questionnaire aimed at identifying the use and availability of software services apart from content and metadata search and retrieval. The first question (*Are you developing or extending services, web or standalone, specifically for your pilot?*) addresses the expected developments to be performed for the pilots. The responses provided brief indications regarding the development plans for each pilot, which are described in full detail in D4.2 and will not be listed here.

The next question (*Have you identified any Europeana Labs services you could reuse in your pilot?*) encourages pilot development teams to navigate and identify potentially interesting services developed in the Europeana ecosystem, which are available through the Europeana Labs infrastructure. Although it was too early in their development path to consider additions and updates of their applications there was interest in available software and tools, namely:

- The Europeana exhibition template
- Culture Collage developed by Monique Szpak
- Twitter EuropeanaBot
- Europeana for Education (E4E) set of two widgets
- Programme Notes HTML5 player
- Europeana GeoSearch
- Cultured Canvas
- Art Color Bits
- I-Treasures"

### 3.1.4 Operational scenarios and planning

The most important part of the questionnaire, from the Technical Space perspective, addresses the actual operational scenarios envisioned by pilots. This includes questions regarding the type of application developed, the nature of access to resources (content or metadata), the expected usage of services and the user policies regarding access and roles in the applications. In parallel, information was requested regarding planning and scheduling of the pilots' development, a detailed report on which can be found in D4.2

As many pilots include more than one applications or different types of distribution for the same one, the results indicate that most of the pilots (4) will offer both web-based and downloadable applications while one will be only web-based and another only downloadable.

Regarding the type of access to content and metadata, the following list of answers (one pilot did not respond to this question) indicates the range of methods expected to be supported:

- Ingest on content curation (by user)
- Ingest on pilot creation (once), Ingest on content curation (periodical updates by curator), Constant access to repository (live search and retrieval)
- Ingest on pilot creation (once), Ingest for new user (once per user), Ingest on application start
- Constant access to repository (live search and retrieval)
- Constant access to repository (live search and retrieval)

All of the pilots indicated that they are willing to allow users to access metadata and content in the application for download (original and user-composed resources), according to their respective licenses.

In terms of services used in the pilot, either specifically developed for it or reused from other providers (such as those available on Europeana Labs), two of the pilots responded that services will be used offline for the preparation of the application, two responded that services will be online and accessed during the application's operation and the other two that their applications contain both types of service use scenarios.

Finally, the following list shows the responses regarding application users and expected roles (*Does the pilot support user accounts and if so, are they also allowed to add and change information?*):

- Read only
- Read, edit and add
- Read, edit and add
- Read only
- Read, edit and add
- No users

As pilots have developed, WP2 Technical Space has maintained dialogue to understand requirements and will continue with this approach throughout the project.

## 4 TECHNICAL SPACE REQUIREMENTS

This section describes the functional requirements that were identified by the processes, interactions and collected information listed in the previous chapters. The primary focus for the Technical Space is to support the content space, the pilots and the activities of the Innovation Space, but we made sure to take into account relevant activities such as those listed in the previous chapter as cases of potential future pilots and application that will be built using the resulting infrastructure.

### 4.1 FUNCTIONAL REQUIREMENTS

#### 4.1.1 Access to content

The Technical Space should provide storage and access to medium and high quality content for use by web-based applications.

Types of content include image, video, audio and text files in various formats.

Metadata should include the URL pointing to the digital object together with a rights statement to define the conditions for re-use.

The access mechanism must be able to filter according to assigned rights statements.

There is an ongoing discussion for the development of the project's Content Space and specifically on the implementation of different access rules for content according to usage scenarios and identified users. In that sense it is envisioned that access for project partner and developers may follow different licensing strategies for content, which may then be available differently to a wider audience (e.g. through the project's hackathons) or the public. In technical terms, the Technical Space must be able to implement access to content based on rights specifically stated for the purposes of re-use scenarios.

#### 4.1.2 Access to metadata

The Technical Space has to offer support for storing and accessing metadata records.

Any data model will be allowed when ingesting while aggregation and publication can be performed using EDM and available profiles or more domain specific standards such as LIDO.

Record serialization can be CSV, XML or JSON with preview interfaces for raw data and available HTML renderings.

Access types include HTTP upload and download, OAI-PMH protocol and HTTP API for the storage layer of the Technical Space.

#### 4.1.3 Dataset validation, transformation & versioning

This requirement specifies the need for grouping metadata records that refer to content into datasets that can be accessed individually. A dataset corresponds to a set of records that are represented using the same data model.

For supported, formal data models (such as EDM and profiles, domain models like LIDO) the system should offer schema validation services based on the XSD and/or Schematron rules.

The system should offer a mechanism to implement formal crosswalks between metadata schemas.

The system should offer the ability to automatically convert between data models for which there are established, formalized crosswalks (e.g. in XSLT).

Datasets can have different versions that must be stored and accessed individually. New versions of a dataset are mainly produced either by transformations to other data models or through the application of data manipulation services (e.g. enrichment, linking). Other types of dataset transformations that produce new versions include the application of services for ID generation, normalization and dereferencing.

Filtering and merging of datasets should be allowed.

#### **4.1.4 Dataset Indexing and statistics**

The Technical Space has to offer dataset indexing to facilitate querying required for services developed for the Technical Space (e.g. metadata cleaning and enrichment) as well as for accessing the aggregated repository.

Indexes for XML datasets will be configured automatically while manual configuration interfaces to implement interactive queries should be available when possible.

Value statistics for import datasets and their transformed versions should be available.

#### **4.1.5 Aggregate multiple sources**

Content and metadata that is used for pilot development will be sourced by cultural heritage repositories such as Europeana and DPLA, as well as by a wide range of available open repositories of multimedia content. The Technical Space should implement import plugins to facilitate the creation and ingestion of datasets (metadata including URLs pointing to digital objects) from identified sources of interest using their exposed APIs.

The Technical Space will implement an import module based on the Europeana and DPLA APIs. Appropriate transformations, such as DPLA to EDM, will be integrated where available.

The Technical Space will identify available 3<sup>rd</sup> party APIs for services such as Wikimedia Commons, Flickr, YouTube or Vimeo and, will evaluate their integration for the ingestion of datasets.

#### **4.1.6 Identification, authentication and IPR control**

The system must authenticate access by identifying a user according to details of his account. Users must belong to an organization for which access rights are configured and assigned to the whole group.

Organizations can have parent organizations that have full access rights for their sub-organizations data.

The Technical Space uses access rights to control access to specific areas of the repository. This is accomplished by assigning privileges to either allow or deny access to a resource (dataset, content or service). Access rights must be able to represent IPR restrictions according to rights statements for resources.



According to the developments of the project's Content and Innovation Spaces, the Technical Space must be able to implement the access rights as those will be defined to allow proper re-use of content.

#### **4.1.7 Metadata cleaning and enrichment**

The Technical Space should offer services for data manipulation. These are grouped in two sets of functionalities:

- Data cleaning routines:
  - value formatting (e.g. dates),
  - tagging using a filtering mechanism,
  - value mapping using controlled vocabularies (SKOS or terminology lists).
- Data enrichment routines:
  - linking to external SKOS vocabularies
  - linking to external Linked Open Data repositories.

#### **4.1.8 Semantic web serializations & technologies**

The Technical Space must include a semantic web repository for storing and accessing metadata that are represented using RDF vocabularies. The system must support semantic web serializations, namely RDF/XML, N-triples and JSON-LD.

The repository must provide a SPARQL endpoint (with support for SPARQL 1.1) and a programmatic API for access.

WP2 will investigate the use of different indexing techniques for RDF data; using the built-in engine of the repository, the Technical Space should offer the ability to query using materialized views.

The repository must allow semantic web engines, such as reasoning systems to be connected for processing of data.

The transformation between XML serializations and semantic web serializations for datasets that are represented using RDF vocabularies should be performed by the system, transparently to the user.

#### **4.1.9 Multi-purpose access APIs**

The Technical Space must offer APIs for accessing functionalities of:

- The authentication layer
- The content layer (storage and IPR management)
- Metadata (for all versions of datasets)
- Services (for those that are invoked by pilot applications)

## 4.2 NON-FUNCTIONAL REQUIREMENTS

To document and ensure the acceptable qualities of the developed system, a set of non-functional requirements and related metrics should be defined and evaluated. The following minimum set of non-functional requirements should be addressed:

- Performance (response time, throughput, efficient resource usage for specific performance requirements)
- Scalability (number of organizations and users, ease of resource allocation to accommodate changing load)
- Availability and Recoverability (ability to maintain an accepted level of performance over time, recovery from errors)
- Data Security and Integrity
- Usability (efficiency, documentation, ease to learn, satisfying for a target user community)
- Interoperability (use services from and provide services to other systems for Digital Cultural Heritage)

## 5 TECHNICAL SPACE ARCHITECTURE

There are various building blocks for the Technical Space; existing, such as the MINT platform, the Europeana API Console and various 3<sup>rd</sup> party APIs and services and, newly designed and developed such as the semantic publication and repository. The high level architecture of the system is presented in Figure 1 (that is extracted from the Description of Work). It is split in three distinct sets of software tools:

- The Data Infrastructure that includes the storage layers for content and metadata from available sources.
- The Metadata Processing Unit that incorporates available services for the management and manipulation of metadata resources.
- The Access APIs that is the set of interfaces that will be made available to professional users and third party developers for the creation of applications that use and re-use digital cultural heritage resources.

Based on the analysis of requirements outlined in the previous sections, the components are further specified as described in the following subsections. These correspond to:

- The Storage layer for metadata and content.
- The Aggregation Platform for managing and publishing DCH resources.
- The Processing Infrastructure and the Services deployment and integration environment.
- The Content Reuse Framework and the Metadata Processing Unit.
- The Access APIs.

It is expected that the Technical Space will also encompass changing requirements and new evolutions even late in the development process. We have started with a pre-release version (3-4 months earlier than the first deployment milestone) to prepare the delivery of efficient software and allow the early cooperation between development teams. The latter will also guide the evolution of the Technical Space as a reference, interactive infrastructure for the open, cooperative re-use of repositories and technical resources for DCH.

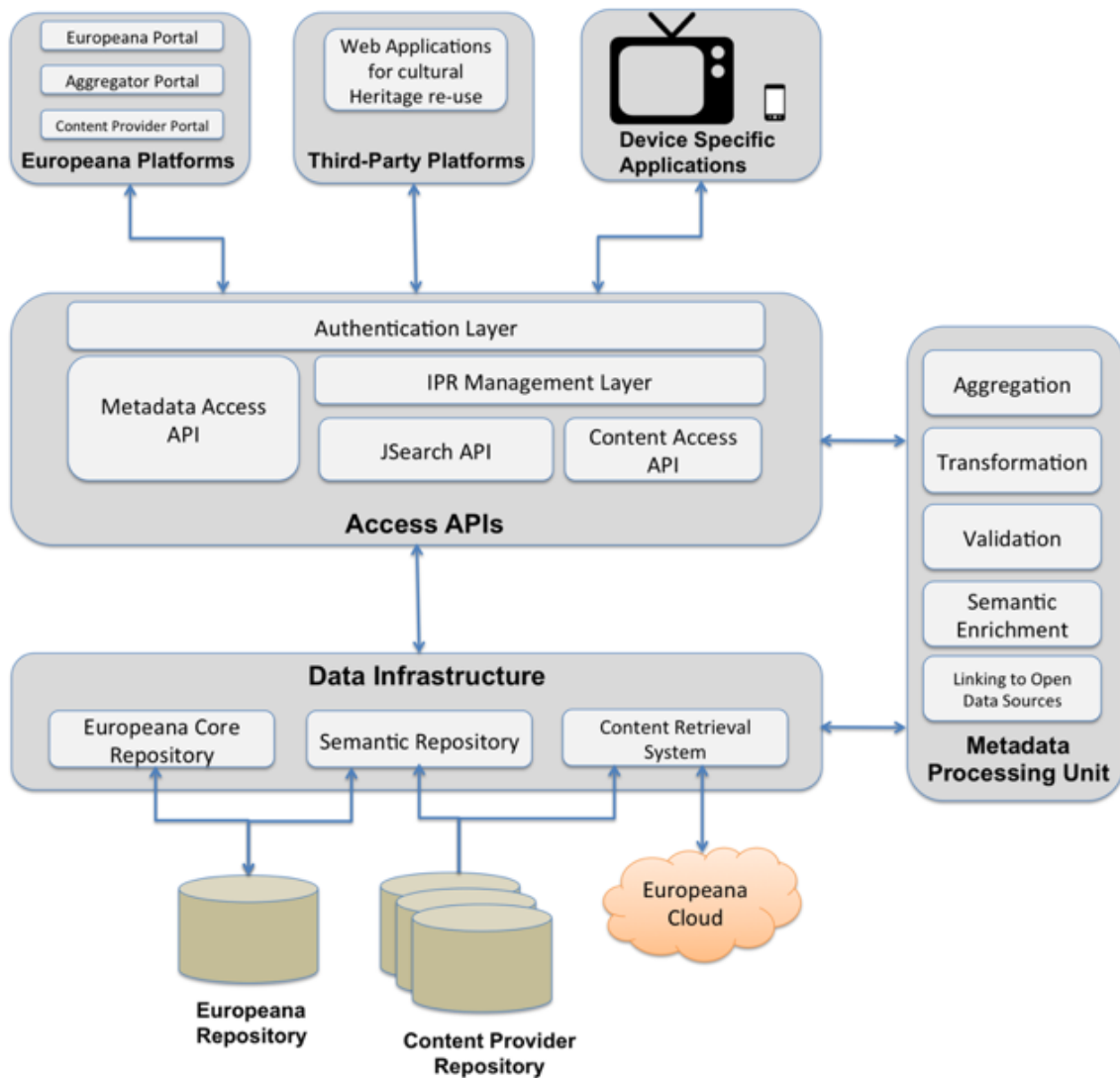


Figure 1. High-level architecture for the Technical Space (from the Description of Work)

## 5.1 STORAGE

The storage layer will interface with the Europeana Repository that holds all metadata for the content ingested to Europeana. It will offer a respective repository in which metadata taken from Europeana or from Content Providers' repositories can be stored, as well as the versions produced after the operations of the Metadata Processing Unit. Published metadata will also be available through a semantic repository, an industrial-strength repository following W3C recommendations for RDF serializations and the SPARQL 1.1 Query Language.

The system will allow for storing, searching, accessing and associating content in an interoperable way with other Europeana initiatives. The Content Retrieval System will provide access to scalable storage services.

### 5.1.1 Metadata

The Technical Space will store all versions of a dataset used in a workflow, along with formal transformations between them. Metadata records and resources are stored in XML-based and Semantic Web serializations.

The MINT backend will be used for metadata storage in the Technical Space. With the introduction of the new workflow data model (see section 5.2), file storage is decoupled from housekeeping data that reside in a relational database. For XML and JSON files, an appropriate NoSQL system will be integrated (currently MongoDB is used for XML files based on which MINT's OAI-PMH server is implemented). MINT will also be upgraded specifically for Europeana Space, with the inclusion of a triplestore for storing metadata that are represented using RDF vocabularies.

### **5.1.2 Content**

For the content layer, the Europeana Content Reuse Framework will be used for the Technical Space. This is a storage and access infrastructure developed by Europeana Creative & Europeana Cloud to allow interaction between content providers and creative industries based on the Extended Europeana Licensing Framework. The latter enables accessing high quality content based on respective rights statements.

### **5.1.3 Importing and Harvesting**

MINT will implement a Data Sources API Layer to facilitate interfacing, importing or harvesting by available DCH repositories. In its first iteration users will be able to combine their imports with resources they locate in Europeana and DPLA, through the integration of their respective APIs. As a next step we will identify available APIs from services such as Vimeo, YouTube or Flickr and implement the required transformations in order to produce metadata records for resources from these providers.

## **5.2 DIGITAL CULTURAL RESOURCES AGGREGATION AND MANAGEMENT**

This is the main platform for aggregating and managing content providers' collections. It is based on the evolution of NTUA's MINT platform according to requirements for the Technical Space and in collaboration with respective initiatives in the Europeana ecosystem.

### **5.2.1 Aggregation and Publication workflow**

The workflow engine allows for the implementing aggregation and publication workflows for all the user scenarios. WP2 leaders are collaborating with Europeana in order to define a common, formal data model for digital cultural heritage workflows. The main characteristics that this model will enable for workflow engines that implement it are:

- Flexibility; the model should be able to represent all identified workflow scenarios in the Europeana ecosystem. These include standard aggregation workflows that aim to publish data to Europeana, workflows with multiple intermediate data models (domain models, serializations for specific services such as previewing and indexing) and multiple publication strategies (Europeana, portals, semantic web endpoints, third-party aggregators).
- Interoperability; the model should clearly describe and enable the use of services developed in the various projects and activities in the Europeana ecosystem.
- Preservation; the model should enable the recording of all required information to document the aggregation and publication activities of content providers.
- Sustainability; the model should reinforce the sustainability of aggregators by allowing the evolution of respective platforms and the transfer of data without information loss.

MINT and Metis, the latter being the new Europeana ingestion system, will implement the data model. These reference implementations will allow the validation and communication of the model to the aggregators and creative industry communities.

### 5.2.2 Processing Infrastructure

The Technical Space will introduce a new Processing Infrastructure implemented in MINT to allow:

- High performance; to cover requirements for constant, parallel access by different applications that are based on creative re-use of cultural heritage content.
- Elasticity; to allow easier resource allocation for all processing-heavy services and accommodate load balancing.
- Efficiency; to perform all the required processing by content providers and application developers using a simple and intuitive processing model.

The Processing Infrastructure will be implemented using Akka (<http://akka.io/>), a toolkit and runtime for building highly concurrent, distributed, and resilient message-driven applications on the JVM. NTUA is also following the specification and development of Europeana Cloud's Data processing service (DPS) that will be used for the transformation of metadata records.

### 5.2.3 Service Integration and Deployment

Using the new workflow model described in section 5.2.1, the Technical Space will enable easier integration and combination of repositories of DCH material and services developed for aggregation and publication workflows. Developer documentation will offer all the required information for enabling services to be used by any workflow engine that implements the data model. The architecture will also allow third-party developers to take advantage of the introduced processing engine as an alternative for the services' computational tasks.

## 5.3 CONTENT REUSE FRAMEWORK

To allow interaction between content providers and creative industries, the Europeana Licensing Framework is currently being extended with a layer that governs access and re-use conditions for the content itself in addition to metadata. The Content Reuse Framework specifies the Content Layer of the Extended European Licensing Framework<sup>1</sup>. It is closely aligned with the work that is undertaken in the Europeana Cloud project on a cloud-based storage infrastructure. NTUA participates in these developments and Europeana Space will be able to use the resulting infrastructure for its content storage and access needs.

NTUA is also developing a suite of tools as part of the CRF, for the extraction of technical metadata and content analysis, called the MediaChecker. The first version that is already available provides a collection of static functions that wrap around some of the best media

---

<sup>1</sup> <http://pro.europeana.eu/web/europeana-creative/extended-europeana-licensing-framework>

analysis libraries available. These include ImageMagick for images, FFMPEG for audio and video and, iTextPDF for PDF files. The second version that is currently being designed introduces a Content Analyzer to perform classification of content and enable the use of more specialized analysis tools.

NTUA is also planning to introduce content-based image analysis tools that can implement respective search engines. In that way users will be able to pose visual queries for specific classes of objects (buildings, people, faces and so on).

#### **5.4 METADATA PROCESSING UNIT**

The Metadata Processing Unit offers content providers the ability to aggregate, validate and transform their metadata to the Europeana Data Model and profiles, or any identified data model of interest. It then allows users to enrich metadata using SKOS terminologies and ontologies as well as to link to external data sources (DBPedia, Freebase, Geonames, VIAF and Getty) obtaining in that way additional information for the objects, places and persons of interest.

The MPU is based on NTUA's MINT platform and its first version is setup<sup>2</sup> for testing by pilots. The current set of functionalities include:

- Import using identified delivery protocols
- Visual mapping editor for crosswalk generation
- Transformation
- Schema Validation
- Data Cleaning
- Reconciliation with SKOS vocabularies
- Publication

MINT is being extended for the MPU in order to include a semantic layer for publication of RDF metadata to the semantic repository described in 6.1. This includes procedures regarding the generation of persistent URLs for all resources in metadata, linking to external sources and validation of the resulting dataset according to Linked Open Data requirements.

The semantic layer enables further processing of metadata using semantic web technologies such as automatic and semi-automatic data enrichment tools. Furthermore, semantic publication enables the introduction of reasoning systems to take advantage of the available ontological knowledge. In particular, prototype semantic query answering systems will be tested in order to extend user queries appropriately by including implied inferences and offer term suggestion and auto-completion functionalities.

#### **5.5 ACCESS APIS**

The Technical Space will provide APIs to the professional users and third party developers for the creation of applications that use and re-use the digital cultural heritage resources.

---

<sup>2</sup> <http://mint-projects.image.ntua.gr/espace>

For content, the APIs of the Content Reuse Framework will be provided, and extended where necessary in order to incorporate new requirements and processes that may be identified by the Content Space of WP3, which investigates rights management for creative exploitation of DCH content. In addition, WP2 works on the definition of an API to establish alignment and interoperability of DCH repositories with the JPSearch framework<sup>3</sup>.

For metadata, the Restful API of MINT will be available for accessing data and services of the platform. Metadata may also be available via the OAI-PMH protocol, through the MPU publication processes, via SPARQL endpoint or the Apache Jena ARQ query engine<sup>4</sup>.

The APIs of the Technical Space will use authentication and implement the access rules for available resources. NTUA will investigate the deployment of APIs for the provision of specific platform's services to developers, such as validation or transformation processes. In a more abstract approach, the processing engine may be available through an API so that application developers may use it to delegate processing tasks for their own services and software.

---

<sup>3</sup> <http://www.jpsearch.org/>

<sup>4</sup> <http://jena.apache.org/documentation/query/>



## 6 PROGRESS ON COLLABORATION TOWARDS EUROPEANA LABS

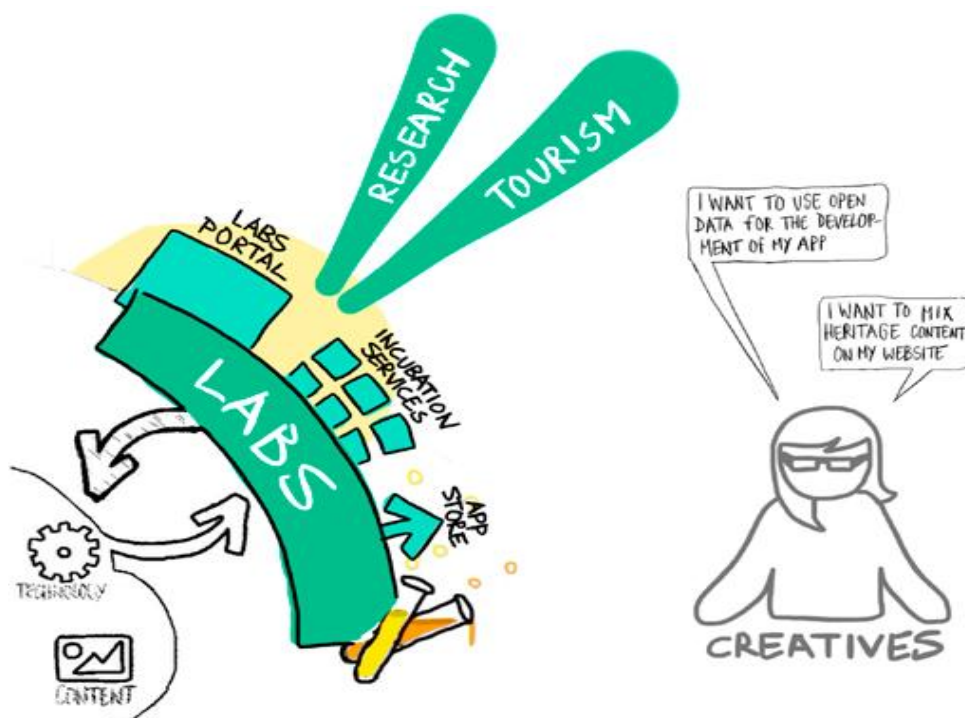
### 6.1 BACKGROUND

In the summer of 2014, the Europeana Foundation released its strategic plan for the period 2015-2020. The plan details the need to increase the number of digitised (higher resolution) items, improve metadata, provide themed collections to engage user input, all from the base of the new platform that will lead to greater interaction and act as a springboard for creative reuse of content. There are three main target groups identified within the plan Professionals, End Users and Creatives; the Europeana Space project predominantly focused on the latter. The relevant extract from the strategic plan and diagram is provided here:

“Creatives

Our third customer group is made up of the developers and entrepreneurs who come up with the new ideas and applications – collectively known as the ‘creative industries’. This part of the economy currently represents 3.3% of the European GDP and we believe Europeana can help to fuel more growth. With ready access to our cultural heritage, this group will use it to develop new services to attract tourists to Europe, inspire new business ideas, create new jobs and find completely new ways to interpret our past. Making the digital material from our memory institutions available to this group in a useful way is difficult and not uncontested. The demands these professionals make are often high.

They need guaranteed stable service levels and access to re-usable, high quality content. They long for collaborative platforms that give them greater latitude and space to think. That’s why we have developed Europeana Labs as an interface for this customer group. We will continue to evolve and improve our co-creation physical and virtual spaces and increase our incubation capability so that we meet their growing needs and expectations.”



As highlighted above, Europeana Labs is a key plank of the Europeana 2015-2020 strategy; a link can be found from the Europeana Space website home page to <http://labs.europeana.eu/>.

The Europeana Labs home page described itself:

“A playground for remixing and using your cultural and scientific heritage.  
A place for inspiration, innovation and sharing.

This is your code, this is your heritage, these are your labs.”

It is clear that although Europeana Labs has an ambitious end goal, to provide access for ‘creatives’ to content and metadata although, it is still very much in its development stage. At its point of conception, Labs had a phased release timeline of:

- private alpha release in January 2014
- public beta in March 2014
- version 1.0 release in July 2014
- version 2.0 release in January 2015.

This scheduled was agreed in line with the event planning and progress of the Europeana Creative project.

Over the developmental period, further ideals have become apparent. Europeana Labs, always incorporated a number of physical labs, as well as access to the content available online, but there has been further progression to incorporate the advertisement and development of hackathons and other events designed to stimulate the creative process.

## 6.2 RELATIONSHIP

One of the potential sources of digitised content for Europeana Space pilots and hackathons is Europeana Labs. However, the development of the pilots, their need for content and an area of experimentation may not yet be met by Labs, which is still very much in its development stage.

As detailed above, Europeana Labs was initiated and subsequently developed jointly between the Europeana Foundation and the Europeana Creative project. At a very late stage of the Europeana Space negotiation process, Labs was written into the Europeana Space DoW with a view to project personnel joining the steering group and potentially taking the reins after the end of the Europeana Creative project in the summer of 2015.

A Skype meeting was held between with Harry Verwayen (Deputy Director, Foundation), Max Kaiser and Katharina Hollas (Coordinator, Creative) and Marion Doyen, Business Development Manager, Antonella Fresa, Technical Coordinator, and Tim Hammerton, Project Manager, (Space) in the final week of January, before the project started, to explore what this relationship could be and what were the plans and objectives of Labs.

During the lunch break of the kick-off meeting further discussion was held with Harry Verwayen to agree an outline for progression and a tripartite meeting, including Europeana Creative to be held at the Europeana New Projects meeting in March. Immediately before the March meeting, Katharina Hollas shared the Europeana Creative deliverable *D1.2 – Europeana Labs website* to provide helpful background information. The idea was for this to be the first of a number of steering group meetings to jointly develop Labs further.

The reality for the initial joint development of Labs was that, although written into the DoW, it was not so easy to immediately make this happen. The Europeana Creative project was already a year into its life cycle and had worked on the definition and shaping on Labs, in line with its work schedule; the Europeana Space project was in its infancy, still building internal relationships and developing its own infrastructure. It was too soon for Europeana Space to participate, but Europeana Creative had deliverables and deadlines for their workload and therefore initial collaboration proved difficult.

Dialogue was resumed after the summer in the early part of September both before and during the Europeana Projects' Congress where Tim Hammerton, was able to meet with Jill Cousins, Director and Harry Verwayen to discuss the position of Labs. As the Europeana Creative project had continued to develop Labs, the Foundation suggested that it was best for Europeana Creative to continue to lead the project, with the option for Europeana Space to rejoin in 2015.

At the end of the Europeana Projects' Congress, Tim Hammerton was able to meet with James Morley, the recently appointed Creative Industries Community Developer, who oversees the Europeana Labs for the Europeana Foundation, and to discuss what was happening and the plans for the coming months as part of the journey of Labs.

During the first week of August, Alex Stan made a conference presentation on the dance pilot at an event also which also featured James Morley speaking about Europeana Labs. This enabled a discussion to take place, which built upon the specific pilot interaction between James and Rosa Cisneros, which had taken place since early August, and built further understanding between Europeana Space and Europeana Labs.

At the Europeana AGM in October, the Europeana Space representatives, Sarah Whatley, Coordinator, Tim Hammerton, Antonella Fresa, Charlotte Waelde, WP3 Leader and Gregory Markus, WP5 Leader, were all able to join the specific Europeana Labs discussion group and business planning meetings to engage in discussion with James Morley and Max Kaiser to strengthen the relationship.

### **6.3 SOURCES OF DIGITAL CONTENT**

As detailed within this deliverable, each pilot is likely to use content from a range of sources, both within and outside of Europeana. It is also the intention that the results of some, if not all, of the pilots could also be placed into Europeana. Although, figures for aggregation, it is expected that each of the TV pilot content partners will provide access to at least 4,000 items, but there is a commitment for final pilot outcomes to also be accessible through the Labs. The destination of results from other pilots/hackathons will be defined on a case by case basis.

At this stage, Europeana Labs holds over 1 million items, divided into 38 datasets; the majority of these comprise collections of images, with more items added on a monthly basis. Based upon Google analytics for early November 2014, the most viewed; with 3,831 views (almost twice as many as the next viewed page) is an image of David Bowie (provided by NISV). With 1,000 views is Europa in een bibliotheek, (Rijksmuseum), the 19<sup>th</sup> most viewed item, all the way down to 100 views of the painting of the Threatened Swan (Rijksmuseum), the 716<sup>th</sup> most viewed image.

The open content available via Europeana is growing at an increasing rate, but may still not be to the level required to service each of the Europeana Space pilots/hackathons at this stage, in accordance with the timescales of the DoW.

As mentioned above, Rosa Cisneros, of the dance pilot has liaised with James Morley of Labs on a semi regular basis to initially explore what could be used by the dance pilot, but then again more widely to look at further project availability. She has produced the following summary of possible Europeana Lab services that could be reused within the Europeana Space project:

1. Europeana's Culture Collage developed by Monique Szpak:  
<http://labs.europeana.eu/apps/culture-collage/>

Europeana Culture Collage is developed by Monique Szpak (@zenlan) and started as a simple exercise in mashing up a few of the new web technologies, incl. Europeana API and Europeana visual data.

2. The Twitter EuropeanaBot: <http://labs.europeana.eu/apps/twitter-europeanabot/>  
Inspired by the DPLABot, this implementation uses Europeana API to present semi-random images from Europeana collections on Twitter.
3. The Europeana for Education (E4E) set of two widget:  
<http://labs.europeana.eu/apps/europeana4education-e4e/>  
E4E is a set of two widgets that automatically suggest cultural resources from Europeana and learning resources from Open Discovery Space. It is aimed at authors of educational pathways. E4E widgets can be integrated into an authoring environment such as OMEKA to augment the authoring process and to enhance the educational impact of a pathway.
4. Programme Notes HTML5 player: <http://labs.europeana.eu/apps/programmenotes/>  
The team built a simple HTML5 player displaying interactive historical information related to a musical work while the track plays through SoundCloud.
5. Europeana GeoSearch:<http://labs.europeana.eu/apps/europeana-geo-search/>  
Europeana Geo Search app for Android
6. Cultured Canvas:<http://labs.europeana.eu/apps/cultured-canvas/>  
UK Cultured Canvas is an easy way to promote culture to a user's Twitter followers. Based on user's preferences/selected criteria, the prototype generates backgrounds of Europeana content for a user's Twitter account.
7. Art Color Bits: <http://labs.europeana.eu/apps/art-color-bits/>  
Art Color Bits introduces an innovative way of browsing painters' works by colours. The software gets all the paintings of an author available in the Europeana database and extracts the main colors from each artwork. After that, the whole collection is arranged chronologically with special attention to the main colors.
8. ArtSpace: <http://labs.europeana.eu/apps/artspace/>  
ArtSpace promotes access to art in everyday situations i.e. Europeana collections can be made available in public places such as coffee shops, libraries, schools and hotels.

These examples demonstrate the opportunities that are available, but are not necessarily the chosen routes of each of the pilot. They do highlight the direction that Europeana Labs is travelling and the potential that it has to engage with 'creatives' for the reuse of content.

## 6.4 NEXT STEPS

The relationship with Europeana Labs is important for the Europeana Space project. The project is committed to utilising content found through Labs, where this is appropriate to the scope of the pilots/hackathons, as well as making end results available within the datasets of Labs, where possible. As hackathon dates are confirmed within the project, liaison will take place to include details on Labs, to reach 'creatives' and engage them within the content reuse process and the journey toward business development and incubation.

Pilot and Hackathon Coordinators would be encouraged to visit Labs during the hackathon development stage to understand what is readily available for them and participants to access. Participants would also be directed towards Europeana Labs to help them to consider ideas and existing apps; an action that would also increase the awareness of Europeana Labs within the general public and the creative industries.

Now that the project has had time to fully define the pilots, as detailed in D4.2, there is scope to knowledgeably engage with Europeana Creative and the Europeana Foundation to consider options for steering the direction of Labs and these conversations will continue over the coming months, as will the development of Labs in terms of quantity and quality of the content that it can make available.

## 7 PLANNING

There are various building blocks for the Technical Space; existing, such as the MINT platform, the Europeana API Console and various 3<sup>rd</sup> party APIs and services and, newly designed and developed such as the semantic publication and repository. The functionality of the Technical Space will be extended in an agile, iterative process according to the requirements that are constantly informed during the steps of Pilots' production. The process relies on a set of stable and efficient services, the outcome of many of the partner's experience as well as the latest evolutions of Europeana and Digital Cultural Heritage in general.

During the first phase WP2 will ensure the availability and support pilots in the usage of the Europeana platform, the repository of Europeana Space metadata (for resources not in Europeana) and, the repository of Europeana Space content. The later can originally be hosted by NTUA for images and text while, the use of the video servers and repositories available via the TV pilot will be investigated according to each specific case. In parallel, there is an ongoing evaluation - especially important for the longer term - of the availability of cloud storage providers, either via the Europeana ecosystem (Europeana Cloud and LoCloud projects) or by the commercial sector.

WP2 will support the use of 3<sup>rd</sup> party (DPLA, Flickr, Vimeo etc.) sources' APIs and tools. Semantic access using open source platforms will aim at intuitive querying of SPARQL endpoints and potential use of their more processing-heavy inference capabilities.

WP2 has deployed the metadata processing unit, enabling the appropriate, formalized sourcing of records and resources from content providers, Europeana and other identified 3<sup>rd</sup> party repositories and, the eventually required manipulations in terms of data and its various models and serializations. In its first iteration, the Technical Space will be complemented with the Data infrastructure, consisting of the semantic repository and the content retrieval system.

In Month 12 of the project the MPU will be released, accompanied with the respective deliverable D2.2. The Data Infrastructure (deliverable D2.3) and Access APIs are due Month 20.

WP2 includes Task 2.5 for monitoring the requirements set by WP4 and WP5 as well as for the continuous testing of the APIs.

WP2 will continue to link with WP3 Content Space, WP4 Pilots and WP5 Innovation Space to coordinate progress across the project. Collaboration will also continue within the Europeana Tech Group to ensure common development prevents duplication of activity across different projects and also as part of Europeana Space's part of the collaborative work in the development of Europeana Labs.

## **8 SUPPORT**

WP2 will organize technical workshops to introduce, train and familiarize Pilot development teams with the features and usage of the Technical Space. This workshop is planned to take place in Brussels on 9 and 10 February 2014, with details still being finalised at the time of submitting this deliverable.

NTUA will assign each pilot with a member of its team as technical assistant for the duration of development. The aim will be to inform and guide providers in the use of the Technical Space, as well as to contribute to the performance of tasks related to it, such as metadata ingestion or mapping.

Finally, NTUA is working on a large-scale documentation process for the whole infrastructure, starting with the Metadata Processing Unit. This will include documents for the description of the tools together with examples, as well as screencasts for individual operations.

## 9 CONCLUSION

This report described the steps taken for the collection and analysis of requirements for the development of the Technical Space. It discussed functional and non-functional requirements gathered from the coordination and development teams for the project's 6 pilots, as well as from a variety of relevant projects and activities within the Europeana Network.

The deliverable includes a more detailed specification of the Technical Space's architecture while also identifying potential synergies and cooperation with existing and under development systems, in various technical levels such as storage infrastructure or third party services for developers and users. Interoperability with the Europeana ecosystem is highlighted as a significant aspect of the architecture.

The deliverable also reports on non-technical aspects of collaboration with the Europeana Network, such as content, sustainability approaches and communication initiatives.

Finally, we introduced planning and implementation priorities to achieve short and long term goals, as well as the processes and tools to support developers and users of the Technical Space.

The delivery of this report denotes the finalization of the requirement analysis phase for the Technical Space and leads to the definition and implementation of its components by WP2. In close collaboration with WP4, pilots can start using the infrastructure for content sourcing and application development as well as for the specification and preparation of the project's hackathon events. In addition, it provides WP3 and WP5 with the necessary information regarding the technical aspects of data re-use in order to design and implement the Content and Innovation Spaces.