



DELIVERABLE

Project Acronym: Europeana Cloud

Grant Agreement number: 325091

Project Title: Europeana Cloud: Unlocking Europe's Research via The Cloud

D5.7 - Product & Services Requirements for implementing Europeana Cloud Services

Revision: 1 (February 2015)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	P
C	Confidential, only for members of the consortium and the Commission Services	

Authors:

Julia Fallon (Europeana Foundation)
Pavel Kats (Europeana Foundation)
Alastair Dunning (Europeana Foundation / The European Library)
Marcin Werla (PSNC)

Reviewers:

Georg Eckes (Europeana Film Gateway)
Lisette Kalshoven (Kennisland)
Maarten Zeinstra (Kennisland)

<i>Revision</i>	<i>Date</i>	<i>Author</i>	<i>Organisation</i>	<i>Description</i>
0.1	Dec 2014	Julia Fallon Pavel Kats Alastair Dunning Marcin Werla	EF EF EF / TEL PSNC	Initial Version
0.2	Jan 2015	Julia Fallon Pavel Kats Alastair Dunning Marcin Werla	EF EF EF / TEL PSNC	Incorporated review by G. Eckes
0.3	Jan 2015	Julia Fallon Pavel Kats Alastair Dunning Marcin Werla	EF EF EF / TEL PSNC	Incorporated review by Kennisland
1	Feb 2015	Julia Fallon Pavel Kats Alastair Dunning Marcin Werla	EF EF EF / TEL PSNC	Final Version

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**This deliverable contains original unpublished work except where clearly indicated otherwise.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

D5.7 Product & Service Requirements for implementing Europeana Cloud Services

[Introduction](#)

[Methodology](#)

[Relationship to Business Model \(Deliverable 5.6\) and Europeana Cloud Services](#)

[Dependences between Product Requirements](#)

[Vocabulary](#)

[Functional Requirements - within scope of Europeana Cloud Project](#)

[PR1 Access to Data Storage for download and upload of metadata and content, via REST API](#)

[Storage of Metadata and Content Records](#)

[Access via API](#)

[Record Identification](#)

[Creation of Datasets](#)

[Access to Technical Metadata and Data Lookup Mechanisms](#)

[PR2 User Authentication: Access permissions and terms of reuse](#)

[User API Authentication](#)

[Structured Access permissions for each data record](#)

[Where an access permission is not specified by the User it shall inherit a default access permission which will be determined by the policy framework developed in PR5.](#)

[Structured method of communicating terms of reuse for each data record](#)

[PR3 Data Services: supporting the data processing needed in a standard aggregation workflow](#)

[PR4 User-Friendly Tools for Access to Data Storage for download and upload of metadata and content](#)

[Harvesting Tool](#)

[Non-Functional Requirements](#)

[PR5 Policy based framework that enable continuity of existing standards for sharing data within Cultural Heritage](#)

[Extension to the Licensing Framework](#)

[Standards for Sharing of Metadata](#)

[Standards for Sharing of Content](#)

[Requirements Summary Table](#)

[PR1](#)

[PR2](#)

[PR3](#)

[PR4](#)

[PR5](#)

Introduction

This document formulates the requirements for the products and services of Europeana Cloud Services described in the Europeana Cloud Business Model for 2015 v0.1¹. The Europeana Cloud Services are a stand alone storage and data processing service which will be run by Europeana Foundation as part of its remit as a Digital Service Infrastructure² from 2016. The basis of the Europeana Cloud Services is provided by the technical and legal infrastructure developed within the Europeana Cloud *project*, funded by the EU from 2013 to January 2016.

The work undertaken to develop these requirements builds on research undertaken in WP5 and WP2 of the Europeana Cloud project.

This document describes the requirements for the products and services to deliver the technical and legal infrastructure within the Europeana Cloud *project*, as well as the additional requirements³ to deliver the vision laid out in the Europeana Cloud Services Business Model⁴. The requirements describe the scope of the products and services necessary to deliver the three value propositions to Aggregators, as primary stakeholders of Europeana Cloud Services;

1. Making it easier to manage data hosting and storage requirements,
2. Provide access to tools and services to enrich data,
3. Enable the sharing of data with anyone, following the COPE principle (Create Once, Publish Everywhere)

A variety of levels of service will be offered. The Basic Service delivered will allow partners to perform the rudimentary functions that can be expected from a cloud environment (these correspond to Product Requirements 1, 2, 3 and 5)

- The ability to upload and download metadata and content via an API;
- Each partner will be able to control and manage user authentication, access permissions & terms of (re)use of their data;
- Possibility to allow others (other partners or third parties) to enrich / annotate data, while keeping the original data unchanged;
- More efficient ways for delivering cultural heritage data to Europeana than currently is the case;

¹ Deliverable 5.6 Europeana Cloud Services Business Model v1.0 https://www.dropbox.com/sh/0dqfomrdon17d3l/AABWnzUuRnxVdipPsej-YSIAa/ECloud%20D5_6%20Europeana%20Cloud%20Business%20Model.pdf?dl=0

² The Europeana DSI will enable connectivity and interoperability for the Europes' cultural heritage institutions, providing efficient solutions for the accessibility of digital resources of European cultural heritage <http://ec.europa.eu/digital-agenda/en/connecting-europe-facility>

³ This distinction is described in the requirements table.

⁴ See footnote 1

- Infrastructure for supporting the data processing tasks needed as part of a standard aggregation workflow ⁵

The Basic+ service adds an extra layer to the Basic service by offer the identical service but also with user friendly (ie with graphical interface) tools for managing data. This will allow for non technical staff to manage data in the Cloud. PR4 specifies this.

A Premium Service will also include extra tools for the quality control and management of data; this will be fully scoped later in 2015. These tools falls outside the scope of the Europeana Cloud project.

Methodology

This Deliverable is based on a number of workshops, consultations and interviews with potential future partners of the Cloud Services.

Research undertaken within the Europeana Cloud project developed the High Level Principles ⁶ and Minimum Requirements for Europeana Cloud ⁷ in 2013 and under WP2 the development of the Europeana Cloud Data Model (See Annex 1) This work was followed by a series of workshops and meetings⁸ to refine the legal and strategic requirements in 2014.

During this time two of the three aggregating parties to the project Europeana Foundation and The European Library revised their long term strategic plans⁹, alongside the third partner Poznan Supercomputing & Networking Centre (PSNC). The subsequent changes in the strategic plans has led to a necessary review of the products and services that are planned to be delivered in 2016, primarily through work undertaken in the Europeana Cloud project.

In order to efficiently refocus the work undertaken to date within the project, and ensure that it meets the long term requirements of Europeana Cloud Services a lean start-up process¹⁰ was followed. The work was undertaken by a small project team made up of representatives from Europeana, The Europeana Library and PSNC. It established a series of iterative developments to develop a minimum viable product for the Europeana Cloud Services

⁵ For examples, the workflow described in The European Library Standards Handbook, <http://www.europeana-libraries.eu/documents/868553/50dd6233-7780-42fa-8e36-9678a0ee03cd>

⁶ Milestone 26: High Level Principles
<http://pro.europeana.eu/documents/1414567/1861924/Europeana+Cloud+High+Level+Principles+V1+-+May+13?version=1.0>

⁷ Deliverable 5.1: Minimum Requirements for Europeana Cloud
<http://pro.europeana.eu/documents/1414567/2240207/D5.1+Minimum+requirements+for+the+cloud?version=1.1>

⁸ Management of Europeana Cloud Workshop Report:
<https://drive.google.com/file/d/0B62MHIFrzNJCeXF5SWFuUFRJcFU/view>

⁹ Europeana Strategy 2020: <http://strategy2020.europeana.eu>

¹⁰ Lean Start Up Process defined by Eric Ries: <http://theleanstartup.com>

through a series of interviews¹¹ with aggregators who are outside of the project, but potential customers of the services.

Using the Europeana Cloud Services Business Model to define the core product and the work undertaken to date within the project, the project team developed the product requirements. Because the functional requirements of the system are strongly influenced by legal requirements, this document is written to describe both these aspects of Europeana Cloud services. These requirements have been cross checked against the existing Europeana Cloud Data Model and found to continue to be compatible.

In addition, as a primary User of the System and core to the sustainability of the system post-project, integration with Europeana tools and services is essential. In addition to User requirements, Europeana specific requirements have been identified and form part of the principle product for example where references to the Europeana Data Model (EDM) are found.

As the technical and legal framework are further refined under WP2 and WP5, it is expected the requirements described in this document will be updated. Any changes made of this nature will be reflected in the document version control.

Relationship to Business Model (Deliverable 5.6) and Europeana Cloud Services

This deliverable has been written in tandem with Deliverable 5.6: Europeana Cloud Business Model. The Business Model presents the vision of the the Europeana Cloud Services, looking at the customer segments, the benefits of Europeana Cloud for each of those segments, the cost model and outlines the service and products that are required to deliver these benefits.

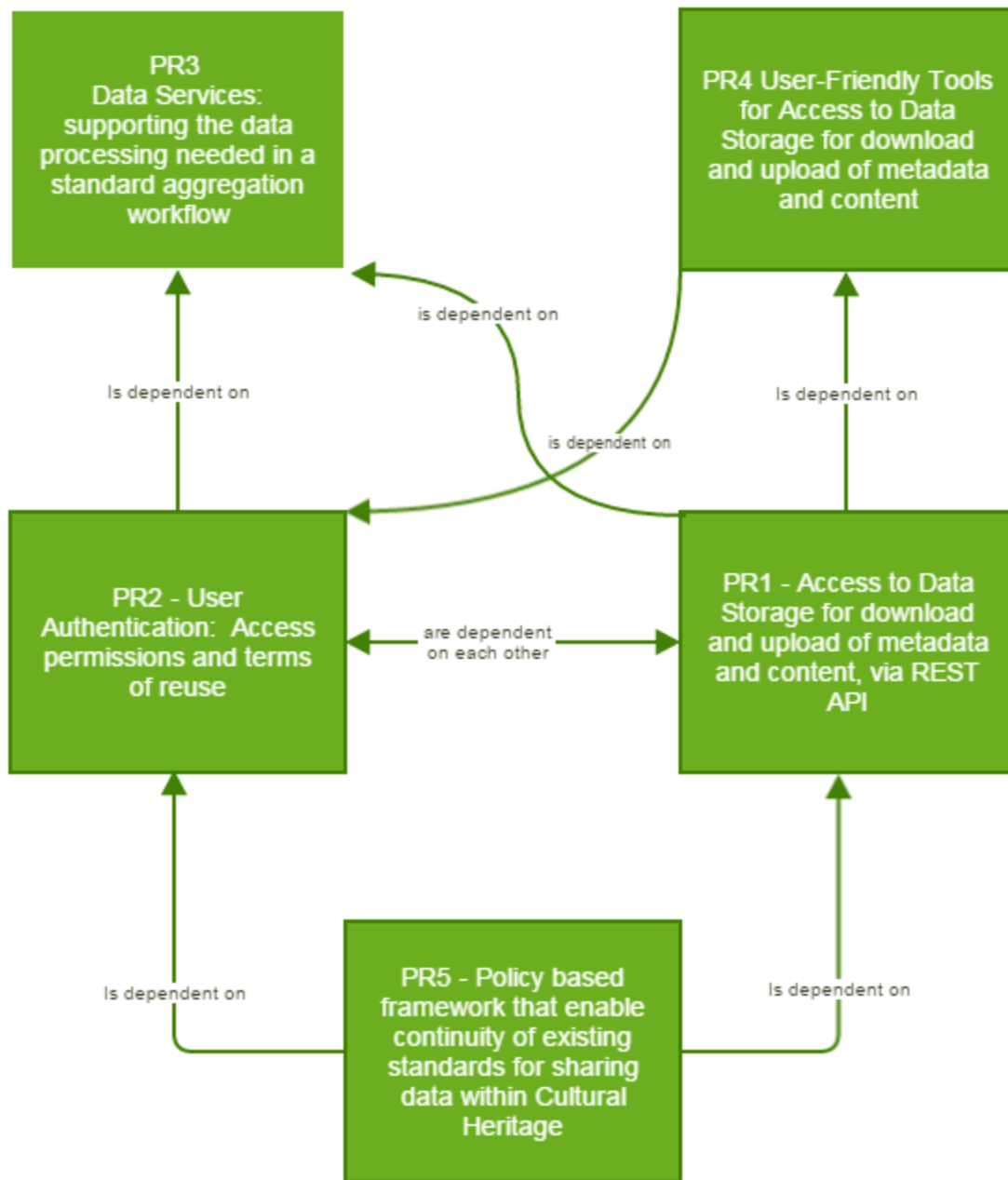
This document takes the products from D5.6 and explore the specific requirements within each of them. It focuses 5 specific products that are required in order to deliver Europeana Cloud for its customer. It does not deal with some of the 'soft' services that will be required - for example, customer helpdesk, or expert support for metadata mapping. This will be dealt with in a separate document in 2015, alongside the development of the governance structure for Europeana Cloud

Dependences between Product Requirements

¹¹ Summary of Aggregator Interviews:

https://docs.google.com/presentation/d/1FEf6rzohE7ozPINFYdd6yi8MUrvAAbM7abKEEaHDHv0/pub?start=false&loop=false&delayms=3000&slide=id.g3bfdddbb_00

This diagram outlines the dependencies between the product requirements in this document



Vocabulary

The list below presents key terms used in the document together with explanations of their meaning in the context of this document, in alphabetical order. These definitions are derived

from the Europeana Glossary¹² and additional definitions¹³ developed within the Europeana Cloud Project.

- **API** - Application Programming Interface, an access point to the system, designed to be used for communication with other software, not intended for regular human use
- **Authentication** - Operation performed by the system, which aim is to confirm the client identity
- **Authorisation** - Operation performed by the system (usually after authentication), which aim is to check if the client is allowed to make specific request to the system (perform specific operation in the system)
- **Client** - A piece of software running on a computer and accessing/using a [service]. Client can be autonomous or operated by human. For example, if the service is a website, the client can be a web browser (operated by human [user]), but also a search engine indexing robot operating autonomously.
- **Content:** A physical or Digital Object that is part of Europe's cultural and/or scientific heritage, typically held by a Data Provider. (Note: You can use the terms Content and Digital Objects interchangeably.)
- **Data** - Metadata or content
- **(Data) Provider** - An organisation that contributes Metadata describing Content that it is offering online.
- **End-User** - anyone who can read or access data stored in the Cloud system ie a member of the public accessing a data record via an API
- **Metadata:** The textual information and hyperlinks that serve to identify, discover, interpret and/or manage Content. Note: This is a general term used to describe any element of Metadata.
- **Record** - Unit of data transferred among [data providers], [aggregators] and [clients]. Can contain [content] and/or [metadata] in many different representations (e.g. different formats).
- **System** - Europeana Cloud system, a set of web services which meets requirements described in this document.
- **User** - human being interacting with the system using client. The User will have an account which defines the roles and permissions they have.

This document follows the ISOs' guidance¹⁴ on how to write standards, using the following clauses to distinguish between required, recommended, permissible and possible features and aspects described as follows:

- **A Requirement;** shall, shall not,

¹² Europeana Glossary <http://pro.europeana.eu/glossary>

¹³ Additional definitions documented D2.2 Europeana Cloud Architectural Design <http://pro.europeana.eu/documents/1414567/2240207/D2.2+Europeana+Cloud+Architectural+Design?version=1.1>

¹⁴ <http://www.iso.org/iso/how-to-write-standards.pdf>

- **A Recommendation:** should, should not
- **A Permission;** may, need not
- **A Possibility;** can, can not

Functional Requirements - within scope of *Europeana Cloud Project*

The Europeana Cloud Requirements are described below.

PR1 Access to Data Storage for download and upload of metadata and content, via REST API

Storage of Metadata and Content Records

The Europeana Cloud Services shall make it possible to store data records that can contain metadata and/or content. The system shall support multiple data formats (especially those common to the Cultural Heritage Industry), because the Users are dealing with different kind of data including proprietary data formats. While data in such formats is not necessarily interesting for the majority of End Users, support for storing it is necessary to ensure Users are able to fully integrate and use the services as a main storage system, and not just a place where data is exported from time to time.

It is not expected that single data records will be very large (e.g tenths of GB/record), but in general from the point of view of User, the system theoretically should be capable of storing any number of data records of unlimited size. There shall be just a common limit related to the total amount of storage space available to the User. It shall be up to the User to decide whether the space will be used to store 10 million small records or 10 very large records or indeed anything in between. Finally there shall be the possibility to store multiple versions of a single data record. This shall provide a history of changes in the data and access previous versions of data, also for the purpose for providing permanent links to data to End Users (e.g. for bibliographic citations).

Access via API

The service shall be available via a machine interface that allows automated interaction - ie via an API. This API shall be highly interoperable, not bounded to any specific technology stack, but based on widely adopted open technical standards. The API shall be designed and implemented to support high performance transfers of data and to process large number of parallel requests. The API shall allow for the storage of and access to single records, but support for batch data upload should also be possible. The API should also provide secure mechanisms for Client authentication and authorisation, on behalf of the User. The communication method chosen for the API access during implementation should ensure

encrypted communication for all operations that modify the data and offer a secure channel as parallel option to an open channel for all data access operations. Thanks to this the Client shall be able to choose the more suitable channel for specific use scenarios.

Record Identification

The system will be dealing with data originating from thousands of Data Providers within Europe, managed by one or many Users . Each of these Data Provider and Users has its own data management system which most probably offers unique identifiers for data records. To make sure that these identifiers do not overlap in any way when data is aggregated by one or more Users, the system (even for internal use) shall generate a globally unique identifier for each provided data record. Additionally Users shall be able to associate such global identifier with one or more local identifiers. Such associations shall be possible to be updated over time. It shall be also possible to interact with the system using both global and local identifiers, assuming that local identifiers will be used always in a context of specific User.

Creation of Datasets

The large number of metadata records in the Europeana ecosystem - tens of millions of data records, continuously growing - requires additional ways to organise data into manageable packages instead of dealing with single data records all the time. Therefore one of the features of Europeana Cloud shall be the possibility to organize data records into sets, ie Users should be able to assign records to datasets. To be able to model the dataflow in the various scenarios for aggregating data Europeana Cloud shall enable each data record to belong to many different datasets, created by different Users - so that each aggregator can organise the same records in different ways as needed.

Access to Technical Metadata and Data Lookup Mechanisms

To make data management, discovery and access easier and more efficient, the system shall automatically record several technical aspects of provided data records, and associated representations and versions (See [Annex 1: Europeana Cloud Data Model](#)).

Such technical metadata shall also contain several pieces of information provided directly by the User. At the time of writing, the list of System Metadata (referred to as Technical Metadata in this document D5.7) generated by the Europeana Cloud System is as follows;

Datasets

- Provider
- Description
- Creation date
- Assignment date (when the dataset was 'owned' by a provider)

- Access Permissions
- Owner User

Versions

- Representation type
- Creation date
- Access Permissions
- Owner User

These lists can be extended in accordance with updated product requirements.

Where possible, the technical metadata elements should have values that are standardised, so there is agreed way to encode the information (i.e. date) or there is a registry of allowed values, or at least a regular expression that the value should conform to.

To ensure this technical metadata is available and usable by the Europeana Network, a connection needs to be made to the Europeana Data Model. An extension to the Europeana Data Model in the form of consideration of new EDM properties¹⁵ would enable this.

Owner	WP2/Europeana, TEL, PSNC
Requirement PR1.1	<p>Storage of Metadata and Content Records</p> <ul style="list-style-type: none"> • The provider shall be able to upload metadata and content records of various formats. • There shall be no limitations on the size of a single record, as well as on the number of records provided by a single User, beside the total storage space available for that User. • There shall be a possibility to store multiple versions of data records.
Requirement PR1.2	<p>Access via API</p> <ul style="list-style-type: none"> • API shall be highly interoperable, not limited to any specific technology stack • API shall support high performance processing of large volumes of requests • API shall allow uploads of single content and metadata records, as well as batch upload • API shall provides mechanisms for authentication and authorised access (see PR2 for details)

¹⁵ Milestone 21 Report with Recommendations for Enhancing EDM
<https://basecamp.com/1768384/projects/2256730/messages/34527075>

	<ul style="list-style-type: none"> ● API shall assure encrypted communication with the system for operations which require authentication, for other operations (especially public access to open data) secure and unsecure (but faster) communication methods should be available
Requirement PR1.3	<p>Record Identification</p> <ul style="list-style-type: none"> ● System shall generate unique IDs for each provided metadata or content record (aka global IDs) ● A User shall additionally register other identifiers coming from the User's system (aka local IDs) ● Client can use local and global IDs to refer to specific record ● Local IDs can be updated
Requirement PR1.4	<p>Creation of Datasets</p> <ul style="list-style-type: none"> ● User shall be able to create datasets in the system and assign data records to them ● Records shall be assigned to one or more datasets, where desirable
Requirement PR1.5	<p>Access to Technical Metadata and Data Lookup Mechanisms</p> <ul style="list-style-type: none"> ● The system shall automatically generate technical metadata for each record and allow User to retrieve using at least these following information: <ul style="list-style-type: none"> ○ global ID ○ local ID ○ type of data representation (e.g. EDM, MARC, JPG, ...) ○ Provider of the data ○ assignment to data set(s) ○ date of upload to the system ○ terms of reuse (see PR2 for details)
Dependencies	<ul style="list-style-type: none"> ● PR1.2: PR2 lists detailed requirements for authorisation. ● PR1.5: PR2 introduces terms of reuse which should be part of technical metadata as well.
Restrictions	No known restrictions
Out of Scope	Any visual end-user interfaces are out of scope of this requirement. It is dedicated to machine interaction.

PR2 User Authentication: Access permissions and terms of reuse

User API Authentication

The Europeana Cloud Services shall be able to authenticate Client and Users access and use of the system via API. It can authenticate End User access and use. A basic level of security should be applied to the system where it is desirable to ensure usage of the system is secure and only by authorised and therefore trusted clients or Users (or End Users). The system shall support the authentication of Clients or Users by providing a valid user name and password.

The system shall enable each User to create a data record, representation or version within that record. For each data record, representation or version that a User creates, they shall have the ability to read, delete or create a new version of that data.

The systems should enable the Data Provider to create a Master or Admin level user account for their organisation. For Users who structure their data according to the Europeana Data Model, this can be based on their EDM organisation profile¹⁶.

The Europeana Cloud system shall be able to authenticate a Users' access to that data in accordance with structured access permissions that are set by the User at the time of upload, or the default permission applied by the system.

Structured Access permissions for each data record

Primary access to the data via an API requires that a User can determine which data records, representations and versions that they have access to, and what permissions they have been granted with regards to that data. As it is likely that a User will upload or edit data on behalf of another User (where for example one User represents an Aggregator and the other is a Data Provider), a User shall have the ability to grant and withdraw access permission to other Users within the system and to view the activity undertaken by Users. The System shall enable access permissions to be granted for each version associated with a data record.

Where a User wishes to update a data record or it's components, a new version is created by the System. The User that creates the new version shall have the ability to grant or withdraw access permissions to the new version.

¹⁶ As suggested in Milestone 21 Report with recommendations for enhancing EDM
https://basecamp.com/1768384/projects/2256730/messages/34527075?enlarge=122688128#attachment_122688128

For each record, representation or version that they create, a User shall be able to grant or withdraw access to another User according to the following permissions;

- Create - where another User can create a record, representation or version on behalf of a User,
- Read - where a User can read the contents of the data record,
- Create a new Version - where a User can read a record or any of its components and they have access in order to create a new version, record or representation (ie not on behalf of the User).
- Delete - where a User can delete a record, representation or version.

For each record, representation or version that they create, or permission has been granted to do so, a User shall be able to grant or withdraw access to End-User according to the following permissions;

- Read - where an End-User can read the contents of the data record,
- Create a new Version - where an End-User can read a record or any of its components and they have access in order to create a new version, record or representation (NB in accordance with the terms of reuse).

Where an access permission is not specified by the User it shall inherit a default access permission which will be determined by the policy framework developed in PR5.

Structured method of communicating terms of reuse for each data record

A User or End User must be provided with clear information about how that data record, representation or version can be used.

To facilitate the communication of this information structured statements that describe the permission granted (or restrictions applied) to the User or End-User shall be implemented.

Where a data record or its components can be read, or a new version can be created, according to the access permissions, it shall be mandatory to assign a reuse statement.

This will be implemented and managed according to policy based framework described in PR5.

Owner	WP5/Kennisland & WP2/Europeana
-------	--------------------------------

Requirement	User Authentication: Access permissions and terms of reuse
PR2.1	<p>User Authentication</p> <ul style="list-style-type: none"> ● System shall support the authentication of Users by providing a valid user name and password, ● A User shall be able to create a data record, representation or version. <ul style="list-style-type: none"> ○ For each record (or component thereof) a User creates they shall be able to read, delete or create a new version.
PR2.2	<p>Structured Access Permission</p> <ul style="list-style-type: none"> ● The system shall automatically authenticate a Users or End Users' access to a data record in accordance with the access permission for that record that is set by the User at the time of upload. ● A User shall be able to to grant or withdraw permission to another User to; <ul style="list-style-type: none"> ○ Create (on grantees behalf) ○ Read ○ Create a New Version ○ Delete ● A User, if the creator or where access permission is granted, shall be able to grant or withdraw permission to End User to; <ul style="list-style-type: none"> ○ Read ○ Create a New Version ● Where not specified at creation, the system shall assign a default access permission.
PR2.3	<p>Structured Reuse permission</p> <ul style="list-style-type: none"> ● The system shall implement a structured method of communicating terms of reuse for each data record ● Where a access to share a record has been granted the User or End User shall specify the terms of reuse ● The reuse information shall be communicated in a structured method using a defined vocabulary
Dependencies	<ul style="list-style-type: none"> ● Authentication Services (D2.2, R5), ● Enhancement of EDM (WP4) ● PR1.5 ● PR5
Restrictions	Alignment with Europeana Data Model.
Out of Scope	The alignment with third party data models.

PR3 Data Services: supporting the data processing needed in a standard aggregation workflow

The platform shall offer a set of core services for building data processing flows involving the data stored in the system (see PR1).

The services shall allow harvesting data into the system from various data sources. The choice of technologies and protocols will be dictated by common needs of the cultural sector. Specifically, since OAI-PMH is a common data exchange standard used by preservation systems, an option to harvest data from OAI-PMH endpoints shall be supported by the system. Other standard choices such as HTTP, HTTTPs and FTP shall be supported. For users wishing to upload content from their personal machines, direct file upload shall also be possible.

It is crucial to provide Users with feedback on the status of the harvesting operations they are running. Especially that today organisations operate increasingly large datasets, statistics are important. These shall be provided by a specialised service which is especially helpful for graphical-user interfaced displaying progress of lengthy operations, such as harvesting. The internal notification service, used by the system shall also notify users about the completion of a harvesting operation.

Each new harvested dataset shall be assigned default access permissions. However, the owner shall also be able to change them.

Transforming data is at the core of aggregation flows which mediate between systems with various standards and data organisation principles. Some aggregators use standard technologies and protocols for their transformation services, XSLT language being the most prominent example. However, others still employ proprietary transformation procedures. The system shall strike the balance between the two by providing an implementation of the most standard choices and letting institutions use the API to build their proprietary flows. Mechanisms available for tracking progress of a lengthy operation, described in the harvesting part, shall also be available for transformations.

Users shall also be able to use these services as well as the basic storage API, described in PR1, to build more advanced flows according to their specific aggregation needs.

The services shall be invoked through the API available in the same way as the API specified in PR1.

Owner	WP2/Europeana
PR3.1	<ul style="list-style-type: none"> ● Data sources and Harvesting <ul style="list-style-type: none"> ○ The system shall support creating data sources using standard data transfer protocols (HTTP(s)/FTP/OAI-PMH) ○ The system shall support scheduling data harvesting from these data sources. Data harvesting will use datasets as an operation unit. ○ The system shall support incremental harvesting when it is supported by a data source ○ The system shall collect statistics about datasets and harvesting operations ○ The system shall support uploading local files and creating datasets from them ○ The system shall define default access permissions to new data which can be overridden by Users
PR3.2	<ul style="list-style-type: none"> ● Transformation <ul style="list-style-type: none"> ○ The system shall support various kinds of transformations of datasets ○ Transformations will create new versions of existing records in a dataset. ○ The system shall support XSLT transformations of XML datasets ○ The system shall support proprietary transformations provided by Users when the underlying technology is supported (e.g. Java-based transformations) ○ The system shall collect statistics about transformation operations
Dependencies	<ul style="list-style-type: none"> ● PR1: PR1 lists detailed requirements for API ● PR2: PR2 lists detailed requirements for authorisation.
Restrictions	
Out of Scope	

PR4 User-Friendly Tools for Access to Data Storage for download and upload of metadata and content

Outside of the original scope of project, but included in response to aggregator needs, PR4 specifies requirements for building End User-facing graphical tools for functionality specified in PR1 and PR2 and available to developers through APIs. The tools will be built on the APIs.

A full specification for this is being built in the European Version 3 project. These will be made available a draft specification is ready at the end of Q1 2015

Harvesting Tool

This tool shall allow harvesting of dataset records into the storage system. This tool will support harvesting via OAI-PMH/FTP/HTTP and direct File Upload. Advanced functionality will include incremental harvesting and scheduled updates.

Mapping Tool

This tool shall feature a visual mapping editor to create mappings applied to datasets during transformations.

Transformation Tool

This tool transforms a dataset using a mapping created with the Mapping Tool or directly using the API.

Validation Tool

This tool allows structural validation of datasets according to a range of formats and validation rules.

Europeana is transforming its own ingestion toolset into a suite of loosely decoupled ingestion workflows. Some of them will also be available for usage by data partners in order to improve their datasets prior to submission to Europeana. These workflows will use Europeana Cloud as the storage layer and processing engines for some of the operations.

Owner	Europeana
Requirements	The requirements for this product are a tool that allows for <ul style="list-style-type: none">● Harvesting data from data providers● Defining conceptual mappings● Transforming datasets based on user-defined mappings● Performing structural validations of datasets

PR4.1	A user-facing tool for harvesting datasets in the variety of options specified in the requirement PR3.1
PR4.2	A user-facing tool for creating conceptual mapping between XML formats.
PR4.3	A user-facing tool for transforming XML datasets using previously created mapping.
PR4.4	A user-facing tool for executing syntactic validation of XML-based datasets.
Dependencies	<ul style="list-style-type: none"> • This is dependent on development of METIS in Europeana version 3 • PR 2 and 1
Restrictions	
Out of Scope	

Non-Functional Requirements

PR5 Policy based framework that enable continuity of existing standards for sharing data within Cultural Heritage

Extension to the Licensing Framework

The Europeana Licensing Framework defines standards for the exchange of metadata and sharing of content within the Cultural Heritage domain. It is a requirement of all providers of data to Europeana that they subscribe to the Licensing Framework by signing the Data Exchange Agreement¹⁷. The Europeana Cloud Service manages both metadata and content as data records, and differently from Europeana’s metadata aggregation neither distinguishes between the two, nor is aware of their format.

As operators of both the Licensing Framework and Europeana Cloud Services, Europeana wish to ensure the continuity of current standards for the exchange of metadata and content by developing an extension to the Licensing Framework. Data Providers (Users) who use Europeana Cloud Services may wish to share their data with Users or End Users of the system but may not wish to share their entire data with Europeana.

¹⁷ Legal Requirements for submitting data to Europeana <http://pro.europeana.eu/licensing>

Users who make use of Europeana Cloud Services can be required to sign an agreement that facilitates the storage and sharing of data records in accordance with the extensions to the Licensing Framework. This agreement can be in addition, or part of the Service Agreement that will govern participation in Cloud Services.

The extension to the Licensing Framework should include;

- Identifying the elements within the current Licensing Framework that also can apply equally to the provision of Europeana Cloud Services,
- The basis of a the legal agreement necessary that can enable the exchange of metadata and content within the Europeana Cloud Services,
- Defining standards or best practices for sharing metadata and content within Europeana Cloud Services.

Standards for Sharing of Metadata

The Europeana Licensing Framework sets the standard that all metadata shared with Europeana is made available via Creative Commons Public Domain Dedication (CC0)¹⁸. This standard is important to the Cultural Heritage Institutions and Data Providers to ensure that their data is shared and reused by creatives, educators and the wider public with no restrictions. However, If applied to all metadata that is stored via Europeana Cloud services this standard is very likely to become a barrier to reuse.

Technical metadata enables the discovery of metadata and content stored within Europeana Cloud should be published via CC0.

Metadata stored in the System that relates to the digital object should be shared under the most open terms possible that encourage reuse. Users may have legitimate legal claim under copyright law to restrict access to certain types of metadata, or a User may have strategic reasons for wishing to restrict access to Metadata.

To strike a balance between the needs of Users and End Users of the System, and provides continuity to existing standards for sharing Cultural Heritage Metadata the extended framework shall recommend a set of minimum metadata fields for which it is mandatory to publish under CC0. The framework will identify common data models used by Users, and will seek to ensure that the metadata fields identified are compatible (or at least comparable).

Standards for Sharing of Content

The Europeana Licensing Framework requires Data Providers (Users) submit a right statement via a dedicated metadata field (edm:rights) for each web resource

¹⁸ <https://creativecommons.org/publicdomain/zero/1.0/>

(edm:resource/edm:isshownat/edm:isshownby) provided. This rights statements describes the copyright status of the object and is intended to facilitate the sharing and reuse of digital objects.

The System (PR2) requires where the access permission enables reading or creation of a new version of the Data Record, representation or version, that it is mandatory to also share the terms on which that use is allowed. This information shall be described in a structured manner using a defined vocabulary. The statements shall be developed according to the standards used within the cultural heritage domain such as the tools and licences developed by Creative Commons¹⁹.

The extended framework should provide the principles and features for a structured system of sharing term of reuse within Europeana Cloud Services which are;

- Consistent with the principles established in the Europeana Licensing Framework,
- Enablers of interoperability between data models,
- Encourages reuse via open licences.

Owner	WP5/Kennisland & Europeana
Requirement	Policy based framework that enable continuity of existing standards for sharing data within Cultural Heritage
PR5.1	<p>Extension to the Licensing Framework</p> <ul style="list-style-type: none"> ● Identifying the elements within the current Licensing Framework that also can apply equally to the provision of Europeana Cloud Services, ● The basis of a the legal agreement that can enable the exchange of metadata and content within the Europeana Cloud Services, ● Defining standards or best practices for sharing metadata and content within Europeana Cloud Services.
PR5.2	<p>Standards for sharing metadata</p> <ul style="list-style-type: none"> ● Technical metadata should be published under open terms <ul style="list-style-type: none"> ○ Identify minimum technical metadata fields to be published via CC0 ● Minimum metadata mandatory to share via CC0
PR5.3	<p>Standard for sharing content</p> <ul style="list-style-type: none"> ● Where sharing rights are granted to the User or End User, a structured method of communicating the terms is implemented,

¹⁹ <http://creativecommons.org>

	<ul style="list-style-type: none"> • The Principles are consistent with the Europeana Licensing Framework • Facilitates interoperability between data models. • Encourages reuse via open licences.
Dependencies	
Restrictions	
Out of Scope	

Requirements Summary Table

Reference	Name	Within Project Scope	Additional Requirements for Cloud Services
PR1	<u>Access to Data Storage for download and upload of metadata and content, via REST API</u>	X	
PR1.1	Storage of Metadata and Content Records	X	
PR1.2	Access via API	X	
PR1.3	Record Identification	X	
PR1.4	Creation of Datasets	X	
PR1.5	Access to Technical Metadata and Data Lookup Mechanisms	X	
PR2	<u>User Authentication: Access permissions and terms of reuse</u>	X	
PR2.1	User Authentication	X	
PR2.2	Structured Access Permissions	X	
PR2.3	Structured ReUse Permissions	X	

PR3	<u>Infrastructure for supporting the data processing tasks needed as part of the standard aggregation workflow</u>	X	
PR3.1	Data sourcing & Harvesting	X	
PR3.2	Mapping	X	
PR3.3	Transformation	X	
PR4	<u>Access to data storage for upload of metadata and content, via interactive, user-friendly tool</u>		X
PR4.1	A user-facing tool for harvesting datasets in the variety of options specified in the requirement PR3.1		X
PR4.2	A user-facing tool for creating conceptual mapping between XML formats.		X
PR4.3	A user-facing tool for transforming XML datasets using previously created mapping.		X
PR4.4	A user-facing tool for executing syntactic validation of XML-based datasets.		X
PR5	<u>Policy frameworks that enable continuity of existing standards for sharing data within Cultural Heritage</u>	X	
PR5.1	Extensions to Licensing Framework	X	
PR5.2	Standards for sharing metadata	X	
PR5.3	Standards for sharing content	X	

Annex 1: Europeana Cloud Data Model

The Europeana Cloud Data Model, developed by WP2 during the scoping of technical requirements in conjunction with the preliminary requirement developed by WP5.

