



Europeana DSI 2– Access to Digital Resources of European Heritage

MILESTONE

MS1.1: INGESTION WORKFLOWS BUSINESS REQUIREMENTS UPDATE

Revision	Version 2
Date of submission	30th of January 2017
Author(s)	Cécile Devarenne, EF
Dissemination Level	Public



Co-financed by the European Union
Connecting Europe Facility

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
2	29th of Jan. 2017	Cécile Devarenne	EF	Amended version after internal and external review; all feedback incorporated from Valentine Charles, Pierre-Edouard Barrault, Marjolein de Vos, Henning Scholz, Yorgos Mamakis, David Haskiya
1	17th of Jan. 2017	Cécile Devarenne	EF	First draft

Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise.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

The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.

Table of Contents

Introduction	4
Definitions and terminology	4
Metis application (general)	5
Metis data model: users, datasets, and organizations	6
User management for the Metis application	6
Dataset and organization management for the Metis application	8
Dataset and organization public API	9
Workflow	10
General	10
Import	11
Transform	12
Process	14
Publish	15
Monitoring the process from the Metis application	16
Metis services	18
Validation service	18
Collections preview service	19
Further prototyping	19
Additional services	20
Non-functional requirements	20
Performance	20
Migration requirements	20
Post release (V1) product evaluation	21
User satisfaction	21
Measuring the success of the product	21
Conclusion	22
Additional documents on Metis	22

Introduction

In January 2016, as part of the effort to move towards a shared aggregation infrastructure, the development of the Metis product has started. As stated in D1.1 - Work and implementation plan to innovate the aggregation infrastructure¹ (further referred to in this document as D1.1), Metis “can be defined as follows:

- Metis consists of several back-end services and a client (that together make up the product).
- The user interface of Metis offers services for harvesting, analysing, transforming, enriching and publishing data on the Europeana platforms.
- The first iteration of Metis targets the Europeana Foundation and The European Library data officers² as its users. Some back-end services developed in the frame of Metis may be used by others; at a later stage Metis can be developed further and offered as a package of tools to expert hubs.
- Most of the services used by Metis are developed using cloud technologies; data processed in Metis is stored in the Europeana Cloud’s storage layer. The Metis client itself is not necessarily hosted in the Europeana Cloud - it is one of its clients.”

To inform the line of work and support the implementation, user research, which had started in the first year of the Europeana Digital Service Infrastructure (DSI), was pursued and detailed requirements identified. User interviews as well as workshops were organized to gather full input and create designs for the Metis client. As Metis is being developed in DSI-2 for Europeana and The European Library users primarily, the requirements presented in this document are mainly reflecting their use cases; these are presented in tables in the chapters below.

In the meantime, consultations with DSI aggregating partners, also started in the first year of DSI, have kept going forward. Requirements expressed during these consultations were added below the main requirements tables in the relevant sections.

Definitions and terminology

Throughout this document, references are made to the **Technical design plan for Metis**; the latest and final version of the document is Version 3³, published at the end of January 2017. All references to this document are indicated as footnotes, with the specific chapter in which the information referred to can be found.

Metis application: client application for use within Europeana Foundation (primarily) and others (after system stabilization), where the functionality of Metis is exposed.

Metis user: user registered in the Metis application. Each user has specific permissions in the application, formalized as “roles”⁴. User requirements as described in this document are requirements for:

1

http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI/Deliverables/europeana-dsi-d1.1-work-and-implementation-plan-to-innovate-the-aggregation-infrastructure.pdf

² The European Library ceased its activities at the end of 2016, but the requirements presented below are still the result of consultation with the library domain as they were gathered in the course of 2016: functionalities which were specific to the previous The European Library ingestion workflows were deprecated while common requirements are presented in this document

³ <https://docs.google.com/document/d/1zIOMDsrB1TTBtomdrzpkmOqmg4t9nMXc7mk6ryuUo-Y/edit>

⁴ See Technical design plan, User roles for the Metis client

- Europeana users: Europeana data officer, Europeana admin.
- Provider user: an expert hub or aggregator data officer.
- Data provider user: a registered user from an institution providing data to Europeana.

Metis dataset or Europeana dataset: an administrative entity consisting of a group of records representing cultural heritage objects (Provided Cultural Heritage Objects (CHOs)). A dataset in Metis represents the entity to which data processing services are applied⁵.

Metis organization: any organization that has a role in the contribution of data to Europeana: data provider (institution that owns the data), provider (aggregator or expert hub), or Europeana itself⁶.

Metis release V1: first release of the Metis application in production. At this stage, it is expected that the current set of tools can be deprecated.

Requirements terminology:

All requirements presented are prioritized according to the MoSCoW method⁷.

M (Must have): requirements labelled as Must have are critical to the current delivery timebox in order for it to be a success. If even one Must have requirement is not included, the project delivery should be considered a failure (note: requirements can be downgraded from Must have, by agreement with all relevant stakeholders; for example, when new requirements are deemed more important).

S (Should have): requirements labelled as Should have are important but not necessary for delivery in the current delivery timebox. While 'Should have' requirements can be as important as 'Must have', they are often not as time-critical or there may be another way to satisfy the requirement, so that it can be held back until a future delivery timebox.

C (Could have): requirements labelled as Could have are desirable but not necessary, and could improve user experience or customer satisfaction for little development cost. These will typically be included if time and resources permit.

Metis application (general)

The following requirements express overall objectives as stated in the product plan⁸ and creative brief.

User requirement	M/S/C	V1	Notes
The product must focus on improving data officers' work environment - and therefore productivity: "tools have to be human centric as an improvement of the current ones"	M	x	User friendliness of the product will be evaluated, see final chapter
The Metis application is a central access point for data ingestion into Europeana; all data operations can be triggered and	M	x	

⁵ See Technical design plan, Metis data model.

⁶ Id.

⁷ https://en.wikipedia.org/wiki/MoSCoW_method

⁸ <https://docs.google.com/document/d/1N2vxxkcUrf71OOUaluJw4BYXZYbfbKXoKJSULHMcjV8/edit>

monitored within the Metis application (as opposed to the current systems where separate standalone tools are in use)			
The Metis application and its back-end services enable an increased focus on data quality	M	x	
The Metis application enables Provider users and Data provider users to monitor the status of their datasets with regards to their publication in Europeana	M	x	
The Metis application enables some Provider users to process and publish data in Europeana	S		This requirement builds on a future use case for the Metis application ⁹ , where providers process their data directly in Metis (case by case request and evaluation); this extension of the users group for Metis is foreseen to happen from V2 onwards

Metis data model: users, datasets, and organizations

As stated in the [Definitions section](#) above, the Metis data model consists of the following three entities: users, datasets, and organizations. Up-to-date detailed information on this model is maintained in one master file¹⁰.

User management for the Metis application

User requirement	M/S/C	V1	Notes
<i>User registration: As a Metis user....</i>			
I want to be able to create my user account in Metis; I want to use my email as a login; I want to enter minimal personal information for the creation of my account	M	x	
I want to be able to indicate whether I am a member of the Europeana Network Association so that the basic contact information registered in the Europeana CRM is prefilled when I fill in my user profile in Metis; I want to be able to register to the Europeana Network Association if I am not a member already when I register to Metis	C	x	Members of the Europeana Network are managed in Europeana CRM, connecting them to the users of Metis would add value but is not considered as

⁹ See Technical design plan, use case 5

¹⁰

https://docs.google.com/spreadsheets/d/1jrL_ie5NU7JnnlQ-kmv8gK_HbGMPMJ1gcrZ_a3y5WhQ/edit#gid=0

			a Must Have for the first release
I want to be able to indicate which organization(s) I belong to when I create my account ; I want to choose my organization(s) among the organizations that are registered in the Europeana CRM as content provider, aggregator, direct provider or Europeana	M	x	
I want to be notified that the organization affiliation(s) I requested is/are pending validation	M	x	
<i>User request approval: As a Metis Europeana admin user...</i>			
I want to be notified that a new user has requested an organization affiliation	M	x	
I want be able to modify the selection of organization(s) requested by the user and select his role for each organization before validating the creation of the account; I want to approve the organization affiliation(s) once the roles have been specified; I want to be able to deny an organization affiliation	M	x	
I want the user to be notified that his affiliation request has been approved or denied	M	x	
<i>User management: As a Metis user...</i>			
I want to be able to view and/or edit my profile	M	x	
I want to be able to request my organization affiliation to be updated	M	x	
<i>User management: As a Metis Europeana admin user...</i>			
I want to retrieve all users registered in the system by searching for their name or searching by organization	M	x	
I want to be able to update all users' information (including organization affiliation and roles)	M	x	
I want to be notified if a Provider or Data Provider user requested an update of his organization affiliation and I want to be able to approve/deny the change	M	x	
I want to be able to deactivate a user	M	x	

Dataset and organization management for the Metis application

User requirement	M/S/C	V1	Notes
<i>Organizations full inventory: identification and deduplication: As a Metis user...</i>			
I want all organizations contributing data to Europeana to be inventoried in the Europeana CRM with a unique identifier and I want to be able to view that information in Metis ¹¹	M	x	
<i>Reference data for organizations</i>			
As a Metis user, I do not want to create organization entries in the Metis application, but I want to have access to the latest updated information about relevant organizations from the CRM (content provider, direct provider, data aggregator)	M	x	
As a Metis Data provider or Provider user, I want to be able to request the creation of an organization in the Europeana systems (Europeana CRM)	S	x	
As a Metis Europeana admin user, I want to be notified when a user has requested the creation of an organization so that I can create or validate the relevant entry in the Europeana CRM	S	x	
As a Metis Europeana data officer user, I want to be able to store technical default information for any Provider organization (e.g.: one OAI-PMH endpoint to be used for all datasets contributed by one Provider organization)	S	x	
<i>Dataset entry creation & update: As a Metis Europeana data officer user...</i>			
I want to be able to create/update dataset entries in the system for any providing organization; I want to fill in the following information: technical information (harvesting details), administrative information (acceptance	M	x	

¹¹ Ultimately, this full inventory of organizations providing to Europeana will not only be used in Metis, but also in Europeana Collections, Europeana Pro and the Statistics Dashboard.

step); I want the identifier of the dataset I am creating to be generated automatically			
I want to link the dataset I create to a Provider organization (edm:provider), and to a Data Provider organization (edm:dataProvider) if the latter is the unique for this dataset, so that Metis users affiliated to those can be notified when relevant	M	x	
I want to be able to create a batch of dataset entries without repeating the information they have in common; I want the technical information entered at the level of the Provider organization to be used as default information for the dataset I create for this organization - if I fill in other values while creating a dataset, these values overwrite the default information filled in at the organization level	C		
<i>Dataset entry creation & update: As a Metis Provider user...</i>			
I want to be able to create/update dataset entries in the system for datasets delivered by my Provider organization	S		This requirement builds on a future use case for the Metis application ¹² , where providers process their data directly in Metis (case by case request and evaluation); this extension of the users group for Metis is foreseen to happen from V2 onwards

In October 2016, following the Aggregator Forum, a joint meeting between Europeana Foundation and the Deutsche Digitale Bibliothek was organized with the following outcomes/requirements:

- There would be added value in implementing a mechanism to feed the Europeana organizations inventory with data from the organizations registry by the Deutsche Digitale Bibliothek.
- There would be added value in aligning the DDB application profile with the EDM profile for organizations¹³ in order to ensure interoperability for future developments.

¹² See Technical design plan, use case 5

¹³ <http://pro.europeana.eu/share-your-data/data-guidelines/edm-profiles>

Dataset and organization public API

As part of the development of Metis, the current Datasets and organizations public API¹⁴ will be refactored with an equivalent level of functionality; the new released API will output data from Metis instead of relying as it does now on Europeana SugarCRM¹⁵. The search and retrieval methods are listed below. Please note that while the calls will remain, the backward compatibility of the API cannot be guaranteed as some field names will be updated to reflect the EDM Organization and Dataset profiles¹⁶. Example responses can be found in the Technical design plan.

User requirement	M/S/C	V1	Notes
List organizations	M	x	
List organizations per country	M	x	
Retrieve an organization	M	x	
List datasets per provider organization	M	x	
Retrieve a dataset	M	x	

Workflow

General

The full workflow for delivering and publishing data to Europeana via Metis consists of several steps which can be grouped in the following categories¹⁷:

- Import: process of getting data from an external system into Europeana Cloud.
- Transform: process of mapping and transforming data from a source schema to EDM Internal, and saving the outcome to Europeana Cloud.
- Process - two types of operations can be considered as part of the process category:
 - Data enrichments (as described in D1.1: process of appending additional data to the delivered data (manually or automatically) and saving the outcome to Europeana Cloud; an example of this is the semantic enrichment step.
 - Additional operations applied to tailor the data to Europeana's systems; an example of this is the hierarchical object generation step.
- Publish: process of checking the results of the applied workflow and publishing the data to Europeana APIs and Collections.

Each step consists of calling a Metis data processing service on a dataset; for the full overview of Metis services, please refer to the Technical design plan (Metis services).

¹⁴ See documentation on Europeana Labs: <http://labs.europeana.eu/api/provider>

¹⁵ <https://www.sugarcrm.com/fr/micro>

¹⁶ <http://pro.europeana.eu/share-your-data/data-guidelines/edm-profiles#Dataset>

¹⁷ Similar categories were also outlined in D1.1

User requirement	M/S/C	V1	Notes
As a Metis Europeana data officer, I want to be able to prioritize a dataset over other datasets, so that when I trigger a process it is executed before the rest of ongoing executions	C		
As a Metis Europeana admin, I want to be able to automate fully the process for “whitelisted” datasets ¹⁸ (import, transformation, process and publication)	S		V1.1, the pilot for implementation will be the User Generated Content produced related to World War I submitted as part of the 1914-1918 thematic collection

Import

User requirement	M/S/C	V1	Notes
<i>As a Metis Europeana data officer user...</i>			
I want to be able to instruct Metis to import datasets to Europeana Cloud using the following protocols: OAI-PMH, FTP, HTTP	M	x	
I want to be able to instruct Metis to import datasets to Europeana Cloud using the other protocols such as IIIF ¹⁹ and/or Sitemaps ²⁰ (see “Report on Linked Data and alternative data acquisition mechanisms by Nuno Freire” ²¹)	C		V1.1
If an error is thrown, I want to be notified, and I want to have access to an error report	M	x	
I want the progress of the ingestion to be shown in the Metis application	M	x	
I want to be notified when an import has completed	M	x	
Once the process is finished, I want the dataset status to be ‘Imported’ and the date of the import to be saved	M	x	
When an import has completed, I want to be able to preview a	M	x	

¹⁸ See Technical design plan, use case 3

¹⁹ <http://iiif.io/>

²⁰ <https://www.sitemaps.org/index.html>

²¹ <https://docs.google.com/document/d/1tLeNWK4ZOaj4n5--nSz0ih4yaxqjsKszS0ctqWjnJhU/edit#>

sample of the imported data			
When an import has completed, I want to know how many new records were added, deleted, updated and I want to know the total amount of records present in the dataset after the harvest	M	x	
When a dataset has already been processed into Metis and is configured to be imported via OAI-PMH or IIIF, I want to be able to import it incrementally for updates	M	x	
I want to be able to schedule imports for datasets that go through regular updates; when this is configured, I want the dataset to be imported without my intervention; I want to be notified when a new import has completed	S	x	
I want to be able to measure the differences between consecutive imports: number of added records, deleted records present in a dataset before and after an import	C		Technical feasibility needs to be estimated
<i>As a Europeana admin user...</i>			
I want trusted API clients to be able to push data (incrementally) to Metis (see in section above: whitelisted datasets)	S		V1.1

Transform

Most of the data processing services applied to the delivered data are expecting data in EDM Internal as input; the mapping to EDM Internal is therefore the first step to be performed after the data was harvested. For this, a mapping tool with a user interface is being implemented, with a main focus on:

- Standard transformation from EDM External to EDM Internal to be automated as much as possible while still enabling the user to edit the data when necessary and flag issues.
- Custom transformations from any input to EDM Internal using XSLT directly.

In the section below, a **crosswalk** is a template defined in the system for a full mapping between a known input schema and EDM Internal; a crosswalk can be used to generate a **specific mapping** for a dataset: this mapping can either stay unchanged (standard automated transformation) or it can be refined according to the specific needs for a given dataset.

User requirement	M/S/C	V1	Notes
<i>As a Metis Europeana data officer user...</i>			
I want to be able to trigger within the Metis application the mapping workflow	M	x	

I want to be able to create a new mapping for a dataset using an existing crosswalk from EDM External or ESE to EDM Internal	M	x	
I want to be able to create a new mapping for a dataset using an existing crosswalk from LIDO to EDM Internal	C		
I want to be able to create a new mapping for a dataset using an existing crosswalk from MARC to EDM Internal	C		
I want to be able to request the addition of a new crosswalk from an input schema to EDM Internal in the system	S	x	
I want to be able to create a new crosswalk myself from an input schema to EDM Internal in the Metis application	C		
I want to be able to apply the mapping to a new dataset and preview the result in a clear user interface (mapping and mapped field values)	M	x	
While the mapping is being applied (transformation step), I want to monitor the progress	M	x	
When the mapping has been applied, I want to be notified	M	x	
I want to be able to edit from the user interface the mapping using standard XSLT string functions	M	x	
I want field values to be automatically flagged when they are not valid against data quality guidelines and I want to view these flags when I have applied the mapping	S	x	
I want to be able to manually flag values and decide whether the presence of these values in the data should generate a warning or a blocker in the mapping report	C	x	
If I have edited the mapping or flagged values, I want to be able to re-apply the mapping (and monitor the progress, be notified when the task is completed)	M	x	
I want to be able to generate a downloadable report for the mapping workflow where the following information appears: number of invalid items after transformation (full dataset), flags	M	x	
I want to be able to reuse an existing mapping - created by me or by any other Europeana data officer user when I process a dataset as an update	M	x	
When updating a dataset, I want to be able to apply a specific mapping on the updated records only (not the full dataset, incremental process) ; if I have changed the mapping, I want to apply it to the full dataset	M	x	
I want to be able to bypass the creation of a mapping in the user	M	x	

interface using a crosswalk and upload an XSLT created outside the system			
I want to be able to apply the XSLT to the dataset on a sample and preview the results	M	x	
I want to be able to edit the XSLT in the Metis application	S	x	
I want to be able to reuse an XSLT I or any other Europeana data officer user previously uploaded	S	x	
When updating a dataset, I want to be able to apply a specific XSLT on the updated records only (not the full dataset, incremental process) ; if I have changed the XSLT I uploaded, I want to apply it to the full dataset	M	x	
I want to be able to download from the application the invalid records after transformation	M	x	

In November 2016, as part of the DSI partners consultations, a joint meeting between The Europeana Foundation and The European Film Gateway (EFG) was organized with the following outcomes/requirements:

- Ingestion workflows are overall similar; however, specific functionalities designed to enhance data quality would be required for EFG to become a user of Metis after V1 has been released.
- Among these, features were identified that would impact the mapping workflow requirements as presented above: possibility to match field values to a controlled vocabulary managed within the system, possibility to manage custom data in a more sustainable way, possibility to create complex mappings (the two last features will be supported in V1 only if the user is creating an XSLT himself and uploads it to the system).
- In addition to the features mentioned above, a migration requirement was also identified for all mappings created from source data to the EFG schema (and further on to EDM): most of the existing mappings are expressed in Perl and as such cannot be ported to Metis.

Process

The process workflow, as described above, consists of applying several data processing services to a dataset. The extended list of services that can be applied is the following²²:

- Identifier generation and itemization
- Cleaning and normalization
- Dereferencing
- Enrichment
- Hierarchical objects extraction
- Technical metadata extraction and thumbnail generation
- Redirects creation

²² See Technical design plan, Metis services

Depending on the dataset, some or all steps need to be applied to the data. It is envisioned that when a workflow has been configured for a dataset, it can be re-applied to the dataset upon updates in most cases (the exception being the Redirects creation step triggered when identifiers of records have changed upon updates).

In order to ensure that the process remains as simple as possible for the user, a workflow manager is being implemented so that the user can trigger this step in one go.

User requirement	M/S/C	V1	Notes
<i>As a Metis Europeana data officer user...</i>			
I want to be able to configure a Process workflow for a dataset when I process it for the first time	M	x	
I want to be able to instruct Metis to start the Process workflow on my dataset after the data has been transformed to EDM Internal (new or updated dataset)	M	x	
I want to be able to apply the Process workflow incrementally	M	x	
I want to be able to monitor the progress	M	x	
When the Process workflow has completed, I want to be able to get a downloadable report, in which individual reports per processed step are present	M	x	
I want to be able to edit a configured Process workflow and save the result	M	x	
I want to be able to trigger a specific step independently from the configured Process workflow (example: Redirects creation)	M	x	
I want to be notified when a Process workflow has completed	M	x	

Publish

This final step consists of checking the data and publishing the dataset (incrementally or not) to Europeana APIs and Collections if the results are satisfactory. The dataset may or may not go through an acceptance step²³.

User requirement	M/S/C	V1	Notes
<i>As a Metis Europeana data officer user...</i>			

²³ See Technical design plan, Use case 1.3

I want to be able to preview the results of a Process workflow; for this, I want to be able to pick up records and view them as XML data and in a Collections preview environment	M	x	
I want to be able to preview the full dataset in an acceptance environment	M	x	
I want to be able to publish the dataset or, if an acceptance step is registered at the level of the dataset, to notify the Provider user that the dataset is awaiting validation	M	x	
I want to monitor the progress of the publishing step	M	x	
I want to be notified that the dataset has been published	M	x	
If a Metis Provider user denies the publication of a dataset, I want to be notified and access the feedback on the issue	M	x	
<i>As a Metis Provider user...</i>			
If an acceptance step is registered at the level of the dataset, I want to be notified when the dataset is ready for publication; I want to be able to preview the dataset in an acceptance environment; I want to be able to trigger publication or deny it if changes are needed	M	x	
I want to be notified when the dataset is published	M	x	

Monitoring the process from the Metis application

Coming from all the requirements above, it was established that the Metis application would be built around two main screens:

- An **overview dashboard** (home page for a user logged in, also referred to in previous documents as “Task Manager”, “User dashboard”) that will enable users to have a quick grasp of what is being done in Metis and to have access to profile pages, among which the most important ones are
- the **dataset profiles** where all relevant information can be accessed for a given dataset and all executions can be triggered and monitored.

This section aims at summarizing the requirements for both the overview dashboard and the dataset profiles; they informed the work done in December and January to create the designs for these screens²⁴.

User requirement	M/S/C	V1	Notes
<i>Overview dashboard: As a Metis user...</i>			
I want to have access to an overview of all executions ongoing in the system	M	x	
<i>Overview dashboard: As a Metis Data Provider user...</i>			

²⁴ See Technical design, Metis client

I want to be able to view in the Overview dashboard the recent history of executions for my datasets (datasets that were entered in the system with my organization as Data Provider)	S	x	
I want to see the latest notifications I received for my datasets	S	x	
I want to be able to search for my datasets or my user account	S	x	
I want to be able to access the relevant datasets full profiles from the overview dashboard	S	x	
<i>Overview dashboard: As a Metis Europeana data officer user...</i>			
I want to be able to view in the Overview dashboard the recent history of all executions	M	x	
I want to see the latest notifications I received for my datasets (datasets that are assigned to my user account) or my user account	M	x	
I want to be able to search for any dataset, organization, user	M	x	
I want to be able to access datasets full profiles from the overview dashboard	M	x	
I want to be able to have access to Europeana CRM from the Overview dashboard (link)	C	x	
I want to be able to access organizations full profiles from the overview dashboard	M	x	
I want to be able to access users full profiles from the overview dashboard	M	x	
<i>Overview dashboard: As a Metis Provider user...</i>			
I want to be able to view in the Overview dashboard the recent history of executions for my datasets (datasets that were entered in the system with my organization as Provider)	S	x	
I want to see the latest notifications I received for my datasets or my user account	S	x	
I want to be able to search for my datasets	S	x	
I want to be able to access the relevant datasets full profiles from the overview dashboard	S	x	
<i>Dataset profile: As a Metis Europeana data officer user...</i>			
I want to be able to view at first sight the main information about my dataset: name, number of published records, main completeness measure, link to dataset in Collections, etc.	M	x	

I want to be able to access/edit the information stored for my dataset entry (see Dataset and organization management)	M	x	
I want to be able to view the last executions performed on my dataset	M	x	
I want to have access to the full history of executions performed on my dataset	M	x	
I want to access all the relevant information related to data quality assurance: valid/invalid records, flags, completeness measures, image caching reports, tier repartition for all records	M	x	
I want to be able to trigger executions for my dataset in the simplest possible manner (see Workflow)	M	x	
I want to be able to see the notifications for my dataset	M	x	
<i>Dataset profile: As a Metis Data Provider or Provider user...</i>			
I want to be able to view at first sight the main information about my dataset: name, number of published records, main completeness measure, link to dataset in Collections, etc.	S	x	
I want to be able to access the information stored for my dataset entry (see Dataset and organization management)	S	x	
I want to be able to view the last executions performed on my dataset	S	x	
I want to have access to the full history of executions performed on my dataset	S	x	
I want to access all the relevant information related to data quality assurance: valid/invalid records, flags, completeness measures, image caching reports, tier repartition for all records	S	x	
I want to be able to see the notifications for my dataset	S	x	

Metis services

All data processing services used by Metis are described in the Technical design plan. Among them, two services were made available to external partners as alpha version for prototyping in November 2016: the Validation service and the Collections preview service.

Both services can be used by external client applications by invoking their dedicated REST API and no limitation is currently placed upon their use.

Validation service

The initial requirement for the service was as follows:

- The Validation service enables validation of xml record based data against an XSD schema and schematron rules.

- Any XSD schema can be uploaded to the service so that records can be validated against it (in January 2017, both EDM External and EDM Internal are managed).
- The service takes as input one or a group of records and sends as a response a report with the status of each record (valid/invalid).

Following the alpha release, feedback from Knowledge Integration was received and requirements incorporated in the Metis roadmap. A second alpha release is scheduled in March 2017, with the following addition:

- The Validation Client will be modified to take the Metis URL as a parameter to the constructor.

Collections preview service

The initial requirement for the service was as follows:

- The Collections preview service enables previewing EDM xml record based data in a Europeana Collections preview environment, which is built against the last version of the Europeana Collections code.
- The service takes as input one or a group of records (maximum 1,000) and sends as a response a report with a link to the group of records in the preview environment and if the process has failed to generate that link, the error(s) that were thrown.
- If the input data schema is EDM External, it is transformed to EDM Internal using a standard XSLT, it is then validated using the Validation service, identifiers are generated using the Identifiers generation service and the result is stored in the preview environment.
- If the input data schema is EDM Internal, it is validated using the Validation service, identifiers are generated using the Identifiers generation service and the result is stored in the preview environment.

Following the alpha release, feedback from Knowledge Integration was received and requirements incorporated in the Metis roadmap as well as bug fixes. A second alpha release is scheduled in March 2017, with the following additions:

- The preview environment will be purged every 24 hours and the purge date will be sent with the response.
- Possibility to overwrite a record: records sent once to the Collections preview service will be overwritten every time they are sent again to the service, enabling the same record to be modified and previewed as many times as needed.

Further prototyping

In February 2016, as part of the DSI partners consultations, a joint meeting between The Europeana Foundation and MUSEU was organized with the following outcomes/requirements:

- The fact that MUSEU operates following a distributed approach as highlighted in D1.1 was discussed and nuanced; however it was clarified that for now, the Metis application is not designed as a solution where all data providers can be users processing their data by themselves.
- A major aspect in MUSEU workflow being the preparation of the data and its mapping to LIDO, an important part of the workshop dealt with this topic. As things stand, the MINT platform²⁵ being used is considered as a satisfactory tool to support this flow.

²⁵ <http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Wiki>

Following this, The National Technical University of Athens (NTUA) expressed interest in testing both Validation and Collections preview services to include them in the MINT platform; input from this testing will inform further requirements for the services.

Similarly, documentation was shared with the following DSI partners: APEF, CARARE, OpenUp!, and also with DDB.

Additional services

At this stage, no other service has been released for external use (shared services); it is envisioned that there may be use cases for other services to be re-used outside of Metis, for instance the EDM data reporting service, but as things stands the service will be implemented for EDM Internal data exclusively. If a requirement is expressed at a later stage, the service can be modified to meet this requirement (post release V1).

Non-functional requirements

Performance

Requirement	M/S/C	V1	Notes
Other than scheduled maintenance SLA is similar as for any other Europeana service as it will be deployed on similar PaaS systems	M	x	
The system supports at least 10 users processing data simultaneously	M	x	
The system supports at least 20 users processing data simultaneously	S		scalability requirement related to the expected developments from DSI3 onwards
The system is able to process on average 2,000,000 records per month (new and updated records)	M	x	
The system is able to process records at the following speed: 150 records per second ²⁶	M	x	
Metis shared services support X number of systems accessing them simultaneously	M	x	scalability requirement depending on the requests made to use the services

²⁶ The current tools process in average from 80 to 100 records per second (with an exception for the mapping workflow: 400 records per second); the goal set for Metis represents a performance improvement

Migration requirements

Requirement	M/S/C	V1	Notes
Dataset information currently stored in Europeana CRM and Europeana UIM must be migrated to Metis	M	x	
The Europeana dataset must be re-indexed in order to ensure that a completeness measure for each record is populated	M	x	This requirement is also a KPI for Europeana DSI-2 ²⁷
The Europeana dataset must be re-indexed in order to ensure that all edm:dataProvider, edm:provider and edm:intermediateProvider values are replaced by the identifiers of each organization pointing to a foaf:organization instance	M	x	
The Europeana dataset should be re-indexed after the normalization service has been applied to all records ²⁸	C	x	

Post release (V1) product evaluation

User satisfaction

Once Metis is released as V1, user satisfaction will be measured:

- For Europeana users (Europeana data officers), interviews will be scheduled where users are invited to provide extensive feedback on the implementation of all requirements above.
- For users of providing organizations that have requested the creation of an account in the Metis application to follow the status of their collections in Europeana, it is recommended that a survey is circulated and follow up interviews are organized to gather meaningful feedback.
- For users of the Metis services, a survey will be circulated to all partners involved in testing the services prior to their production release.

²⁷ "Have *one* completeness measure in the API output, based on the Metadata Quality Assurance Framework developed by Péter Király"
http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_WG/DataQualityCommittee/DataQualityCommittee-2016Report.pdf

²⁸ See Technical design plan, Metis services

Measuring the success of the product

Once Metis is released and stabilized as a production system, the success of the product must be measured against the main objectives stated in the beginning of the document:

- Productivity of data officers
Currently, between 100 and 300 datasets are processed every month in the Unified Ingestion Manager (UIM); as current, independently from the amount of time spent by a data officer on data quality assurance and the processing time, all datasets take in average 40 minutes to be pushed through the full workflow and a lot of manual intervention is needed. The overall user friendliness of the Metis application will be evaluated against this: time must be freed for data officers to focus exclusively on data quality assurance.
- Increased focus of data quality
As stated above, data quality assurance is core to the system. Features related to data wrangling as well as automated quality reports will be implemented to support this. It is recommended that these features are evaluated against the requirements stated in this document (see [User satisfaction section](#) above); overall data quality of the Europeana dataset must also be improved, which should be measured against various KPIs over time.

Conclusion

The requirements as presented in this document enabled precise scoping of the Metis product as it is developed in DSI 2. Once the product is released as V1 and has gone through a stabilization phase, it can be envisioned that work can start to extend the Metis application to support requirements expressed by external aggregating partners. As of January 2017, EFG was identified as a potential candidate; however, more detailed requirements would still need to be identified as well as the costs of the migration of infrastructure that would imply.

More DSI partner consultations are scheduled already to progress on this task, which will also help in gathering new requirements for the second line of work consisting in building on the shared services so that they can be reused by other aggregating platforms.

Additional documents on Metis

All Metis documentation is gathered and regularly updated here:

<https://app.assembla.com/spaces/europeana-ingestion/wiki>

Timeline (updated in December 2016):

<https://docs.google.com/spreadsheets/d/1371G4zUuZcDIUKnhP7T5ua-jAC-uwriFr46onNKvsP8/edit#gid=911659809>

Roadmap (updated December 2016):

https://docs.google.com/spreadsheets/d/14qWYgwMVLU_A6mIYWoEFrq6em6Ur5sGPcEs32HhBiv8/edit#gid=88511548

MS1.1: INGESTION WORKFLOWS BUSINESS REQUIREMENTS UPDATE

Metis Product plan (September 2016):

<https://docs.google.com/document/d/1N2vxxkcUrf71OOUaluJw4BYXZYbfbKXoKJSULHMcjV8/edit>

Metis FAQ regarding implementation (June 2016):

<https://docs.google.com/document/d/1Hpq0ZmwhJVDH6fAsnt8dIMbFV1CXEEv3vkdmOSJB13M/edit>

Metis backlog:

https://app.assembla.com/spaces/cnzHpQChGr4QFdacwqjQYw/tickets/new_cardwall?default_list_cardwall=milestone:2637283