Functional specification for the Europeana Danube release

Europeana v1.0

Grant Agreement Number: 558001

Deliverable number	*D3.2*
Dissemination level	*Public*
Delivery date	*31 August 2010*
Status	*Final*
Author(s)	*Ruth Bloomberg, Makx Dekkers, Stefan Gradmann, Mats Lindquist, Carlo Meghini, Julie Verleyen*

1   Introduction

This document sets out the functional specifications for the second operational release of Europeana.eu, also known as the Danube release. The aim of the Danube release is to provide sophisticated functionality for Europeana's digital library of over 10 million items of Europe's cultural and scientific heritage.The vision for Europeana in 2011 is that it will be a distributor, facilitator and innovator.

While the 2010 Rhine Programme began the process of meeting these challenges, the Danube Programme will offer richer functionality and so move a step further towards the Europeana vision. Key to achieving this aim is the Europeana Data Model, which is released along with this specification document.

1.1   About this document

Introduction
The functional specifications are part of a living document. This document is structured according to a principled methodology of software and is the outcome of a complex process described more fully in Section 2.

The functional specifications are maintained and edited on a wiki. Each specification is linked on the wiki to:
- the requirement it is designed to meet, and
- the software development activity designed to implement it.

Section 2: Methodology and process
This section outlines the Danube methodology and details the four stages of the development process.

Section 3: EuropeanaLabs
This section briefly outlines the structure of the wiki and explains how to access the Danube requirements, functional specifications and related documentation, all held in EuropeanaLabs.

Section 4: Guide to the requirements and functional specifications
This outlines each of the Danube requirements and the functional specifications associated with it. The section includes links to the relevant wiki page.

Section 5: The Europeana Data Model
This section introduces the Europeana Data Model (EDM). The EDM is the structural backbone of the Europeana data, and is therefore key to several Danube functionalities.

Note: The EDM is described in full in two separate documents:

1. the EDM specification: this document sets out the formal definitions of the classes and properties that make up the model
2. the EDM Primer: an informal introduction to the model that uses examples to illustrate its main classes and properties.

Section 6: Future development
The final section gives an overview of how the future development of Europeana may be determined.

1.2 Note on supplementary documents

There are a series of supplementary documents which give background information and further details to support the Functional Specification.

These documents are contained in a dedicated folder on the Project collaborative platform.

Please refer to (https://version1.europeana.eu/group/europeana-collaboratory/documents) following the path:
Folders » WP3 » Deliverables » D3.2 Functional specifications for Europeana Danube release.

2 Methodology and development process

2.1 Methodology

The Europeana development process has been carried out according to the Rational Unified Process (RUP) methodology, which:
- prescribes iterative and incremental improvement of products and services, and
- is premised on close cooperation between the business and technical teams.

2.2 Process

The process is divided into four phases:

1. Requirements definition and validation
2. Functional and technical specifications
3. Development, and
4. Integration, final testing and implementation into production.
Note: these phases are described in detail in Section 2.3

- Coordination is the responsibility of a Product Manager from the Operations & Tech team and a Product Developer from the Business team
- Requirements ownership is the responsibility of a number of individuals from the Europeana project
- Decisions are made by the Executive Director, Business Development Director and Operations & Technical Director, based on proposals made by the Product Manager and Product Developer.

The graphic below shows the organisation of the requirements and specification process. When this process is complete, final decisions are the responsibility of the Executive Director.

```
┌─────────────────────────┐              ┌─────────────────────────┐
│   Business Director      │              │   Technical Director     │
│   Harry Verwayen         │              │   Jan Molendijk          │
└─────────────────────────┘              └─────────────────────────┘
     │   ┌─────────────────────────┐  ┌─────────────────────────┐   │
     │   │   Product Developer      │  │   Product Manager        │   │
     ├───│   David Haskiya          │  │   Julie Verleyen         │───┤
     │   └─────────────────────────┘  └─────────────────────────┘   │
     │   ┌─────────────────────────┐  ┌─────────────────────────┐   │
     │   │ Product Development Team │  │   Development Team       │   │
     └───│ David, Annette, Georgia, │  │ Borys, Jacob, Serkan,    │───┘
         │ Anne-Marie               │  │ Shamilla, Sjoerd         │
         └─────────────────────────┘  └─────────────────────────┘
```
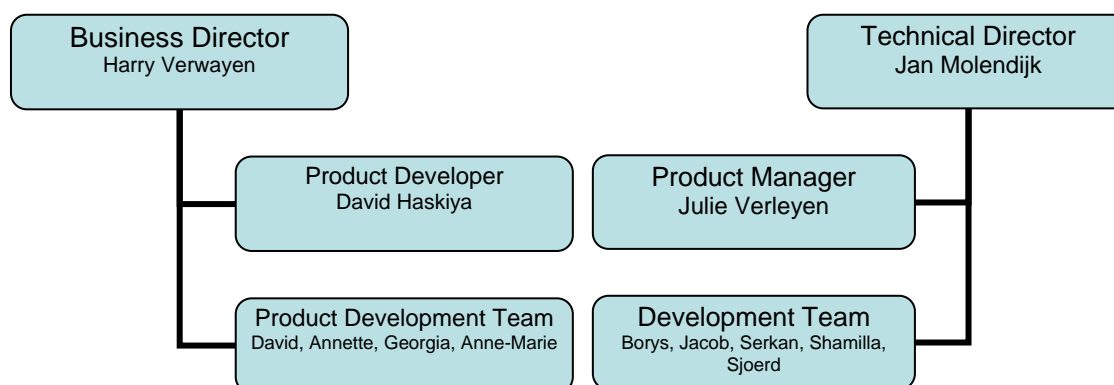
2.3   Phases 1-4

2.3.1   Phase 1: Requirements definition and validation

The development process began with the requirements definition and validation phase.

The outcome of this phase was the development by the Europeana business team of a set of requirements related to the following broad topics:

- Search and retrieve
- Ingestion and the Europeana Data Model
- Reaching the end-user
- Multilingualism, and
- Mobile access.

Stage 1
The first stage in this process was to collect, list and compile brief descriptions of a large number of ideas. These were sifted to eliminate any that were either impossible or undesirable, leaving a list of around 50 potential requirements. For each of these, we produced longer written descriptions, conceptual designs, screenshots of existing similar services and simple mock-ups. **Note**: these visualisations were created in order to give everyone a common understanding of how these ideas might work in practice and the challenges involved in bringing them to reality.

Stage 2
The second stage took the form of a series of validation sessions held between the business team requirement owners and the technical team. On the basis of these discussions, each idea was scored for 1) feasibility and 2) value.

Notes:
- *Feasibility* means both technical and legal feasibility. **Note**: the scores were estimated in the context of the timeframe for Danube. Post-Danube feasibility may change considerably.
- *Value* refers to value for both Europeana's partners and its end-users.

Stage 3
The third stage was to assign each requirement a priority on the following basis:

Priority 1 (High value – High feasibility): Will go live in Danube
Priority 2 (High value – Low feasibility): A prototype will be placed in the Europeana ThoughtLab at the time of Danube and subsequently reevaluated.
Priority 3 (Low value – High feasibility): Wait, refine and reevaluate for post-Danube releases.
Priority 4 (Low value – Low feasibility): Undesirable.

For ease of comparison, the requirements were each formatted into the same template in the EuropeanaLabs wiki. Each was then placed on a horizontal value/vertical feasibility graph, allowing us to gauge their comparative importance.

This prioritisation process was completed and all priority 1 requirements signed off by the business team on 9 July 2010.

2.3.2 Phase 2: Functional and technical specifications

In this phase, the requirements were analysed in detail from a technical perspective. This enabled us to:
- define each requirement in technical terms
- arrive at a technical solution, and
- estimate the development effort.

The technical specification phase was a communal effort involving the entire Europeana technical team under the coordination of task leader Carlo Meghini. As with the rest of the process, this is an iterative process, with the requirements owners from the business team sometimes required to clarify and refine their requirements in collaboration with the technical team.

The completed technical specifications were passed to the business team for final sign off.

The graphic below shows the Danube iterative development process.
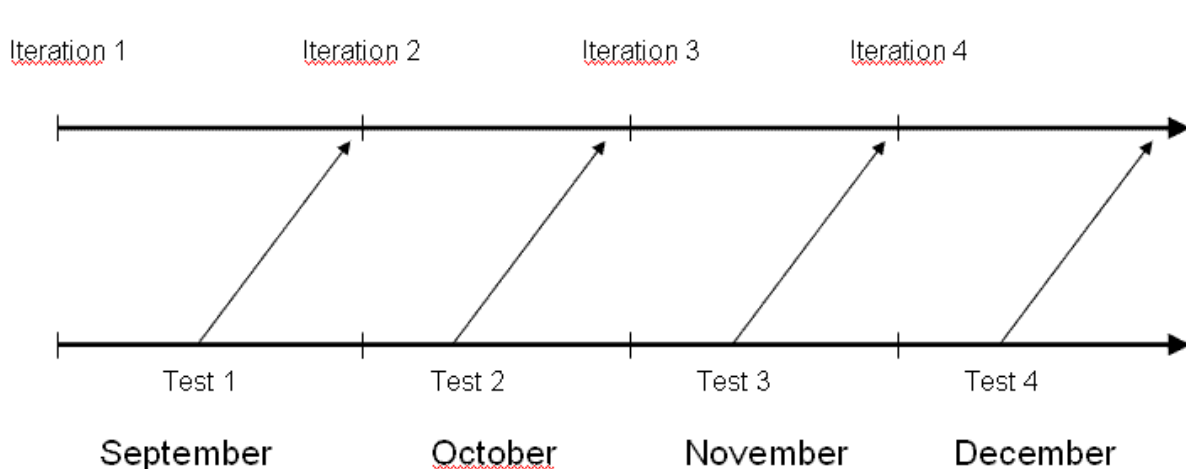
2.3.3 Phase 3: Development

Development begins with the completion of the technical specification phase and a comparison of cost estimates against available resources.

The development of new functionality and features is carried out in parallel with bug fixes and any feature changes to existing functionality. It also involves the testing and integration of software deliverables from Europeana group projects Europeana Connect and ASSETS. In practice, then, this phase is carried out by a distributed team.

The development of the Danube release is planned as follows:

- There will be four monthly iterations in the period September 2010 to January 2011.
- Each iteration will see a number of requirements developed and implemented in the Europeana test version.
- The requirements will go through a preliminary acceptance-testing phase and any resulting change requests fed back into the next iteration.
- Europeana also plans to run usability tests in parallel with the development. This will give us early end-user feedback and the opportunity to take that feedback into account during later iterations.

Features that are considered ready to go live after this testing phase may be included in intermediate (pre-Danube) releases.

2.3.4 Phase 4: Integration, final testing and implementation into production

The aim is for all the requirements to be implemented in the Europeana test version by the end of the fourth iteration. We will then instigate a period of acceptance testing, technical tests and final feature changes before releasing the Danube version of Europeana in April 2011. This period may also be divided into iterations.
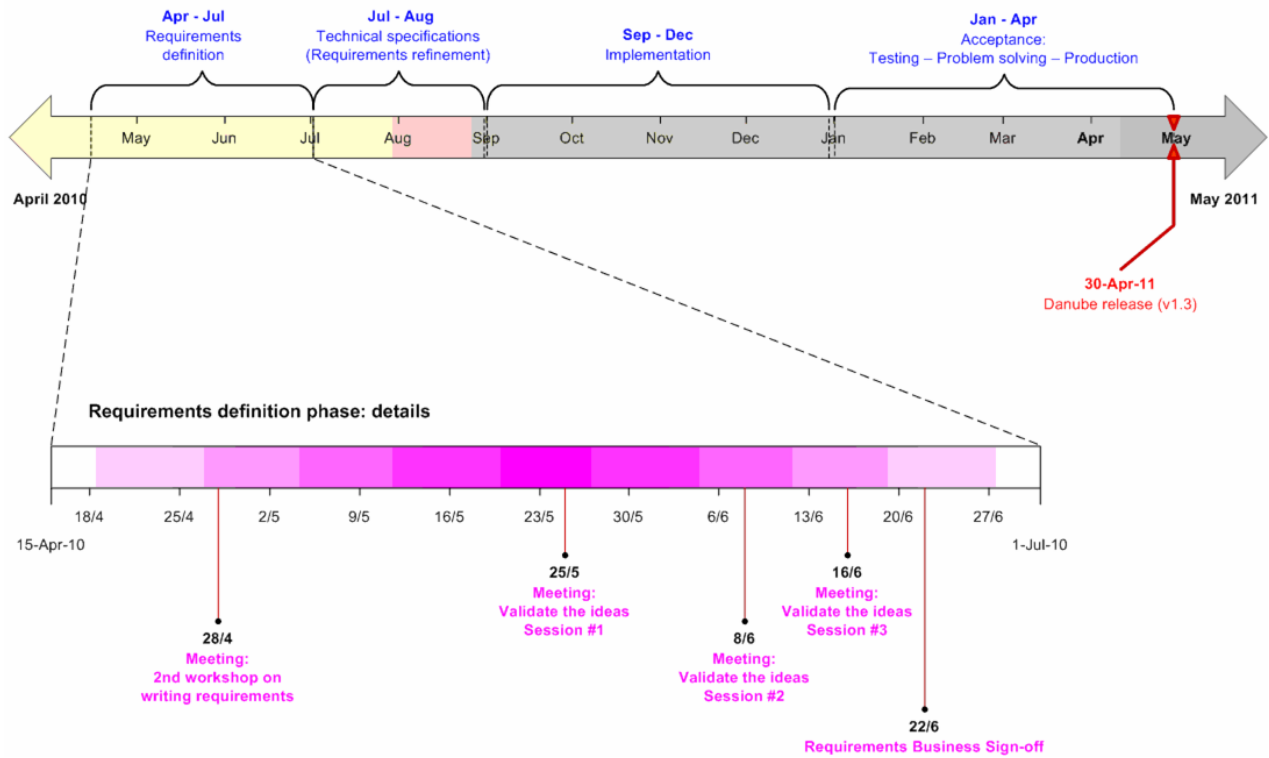
2.3.5 Danube timeline

The graphic below shows the Danube planning process.

DANUBE
Last updated: 2010-06-15

3   EuropeanaLabs

The Europeana requirement and specification process is tracked and documented in EuropeanaLabs, our dedicated web environment for system development and documentation.

Each requirement is identified by a number which corresponds to a ticket in the EuropeanaLabs task tracking system. This makes it simple to follow the progress of each requirement through submission, technical specification, development, testing and implementation into production. The ticket is also the placeholder for all the descriptions and specifications that document the requirement.

EuropeanaLabs can be accessed by Europeana Office users and Europeana group project members. **Note**: read-only access except for users with a log-in.

To view the requirements defined for Danube, refer to
http://europeanalabs.eu/wiki/DanubeRequirements.

4   Guide to the requirements and functional specifications

This section:

- briefly describes the Priority 1 requirements
- provides links to the relevant business requirements and functional requirements/technical specifications.

The requirements are grouped under common themes.

The following requirements have been proposed by the Business at Europeana and will be subject to further technical analysis in the technical specification process before they can be signed off. They are also dependent on development resources.

4.1   Improved search and navigation

4.1.1   Improved ranking of search results

The intention of this requirement is to define and provide:

- improved ranking of search results
- clearer and more transparent display of ranked search results, and
- a user feedback mechanism.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchImprovedRanking
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationImprovedRanking

4.1.2   Hierarchical object display

The general idea of this requirement is to better display complex hierarchical objects in Europeana. Complex hierarchical objects refer to an archival series or sub-series and their underlying individual items. **Example**: for an Eastindiaman's crew manifest:
- the top object represents the manifest as a whole, and

- the sub-objects represent the individual scanned pages of the manifest.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsDisplayHierarchicalObjects
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationDisplayHierarchicalObjects

4.1.3   Map search and display

This requirement aims to provide users with a means of exploring the contents of Europeana along the spatial and the temporal dimensions, or possibly a combination of the two. The idea is to provide the user of the portal with a map- and timeline-based interface with the option to display any search result on a map. Objects ingested into Europeana are connected to specific places and times in various ways; especially interesting are their place of creation, the place they are currently stored in, and the places they are about.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationDanubeRequirementsBrowseSpatial
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationBrowseSpatial

4.1.4   Improved and integrated Timeline

The general idea of this requirement is to allow users to search within timespans and see their search results displayed on an integrated timeline. This feature builds on the timeline released in the Rhine version.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchImprovedTimeline
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationImprovedTimeline

4.1.5   Icons on object for licenses

The aim of this requirement is to give the user clear visual cues as to which license a digital object or Europeana metadata record is available under. Icons will conform to industry or formal standards. **Example**: Creative Commons.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsDisplayLicenses
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationDisplayLicenses

4.1.6   Icons on special content objects

The idea of this requirement is to give the user visual cues that will highlight certain types of content. We want to draw users' attention to content types such as:

- downloadable eBooks **Examples**: ePub, PDF
- high-resolution images (2000x2000 pixels plus)
- user-created objects
- objects with multiple images
- objects of extra high documentation quality **Examples**: narrative texts, stories, popular descriptions
- objects suitable for teaching purposes

- full length films.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsDisplaySpecialcontent
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationsDisplaySpecialcontent

4.1.7 Brief and Detailed Record View Translation Using Google Translate

Search results in Europeana are returned in every Europeana language. This requirement would enable the user to have a Europeana object description translated into their language of choice. This can be done with free services such as Google Translate and Babelfish.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchMultilingualismResultsTranslation
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationSearchMultilingualismResultsTranslation

4.1.8 Facets for licenses

The general idea of this requirement is to make it easy for a user to filter search results based on the license type he/she is looking for. This means it will be easy for users to find, for example, content available under Public Domain or Creative Commons licenses.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchRetrieveLicensefacet
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationSearchRetrieveLicensefacet

4.1.9 Improved Accessibility for Visually Impaired and Blind People

Example:
- A blind person visits Europeana. His computer is equipped with assistive technologies configured to improve his access to online websites. He has a braille reader and a text-to-speech tool.
- When he tries to access Europeana, his assistive mechanisms cannot interpret the extensive visuals. When he tries to read the Results page, he receives the message *Click here for more information* in place of every thumbnail. He needs textual equivalents for all "essential" visual elements of the portal.

The W3C Web has produced extensive Content Accessibility Guidelines 2.0 which will serve as our guidelines for the future design of Europeana. These guidelines should be taken into account in every requirement for Danube and beyond.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsImprovedAccessibilityForVisuallyImpairedPeople
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationImprovedAccessibilityForVisuallyImpairedPeople

4.1.10 Portal design modifications

The Danube release will add a large number of new functions to Europeana. The existing portal design will be modified in order to integrate these functions into the overall layout and navigation.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsDesignPortal
- Functional specification: pending decisions made on all requirements

4.2 Contextualising and enriching content

4.2.1 Improved Virtual Exhibitions

A first Europeana Virtual Exhibition will be released as part of the Rhine release programme. In the Danube program, the objective of Virtual Exhibitions will be to focus on the engagement, involvement and active participation of end users and partners. They will have the opportunity to comment on Virtual Exhibitions, share them in social spaces and create their own galleries. In addition, portions of the Exhibitions can be distributed outside the Europeana portal and exhibition space.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsEndUsersVirtualExhibition
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationImprovedVirtualExhibitions

4.2.2 Improved tagging and tags in search

This requirement enables registered Europeana users to:

- tag objects
- make them visible to other users, and
- find them via the search index.

Its purpose is to enable users to add keywords to Europeana objects in their native languages. In this way, we will harness the energy of the Europeana user community to enrich the object descriptions.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsReachingTheEndUserImprovedTaggingAndTagsInSearch
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationReachingTheEndUserImprovedTaggingAndTagsInSearch

4.2.3 Social media streams on portal

Europeana is increasingly using social media such as Twitter and Facebook to market the project and portal and to engage with our user community. The social media streams will show these activities in real time.
- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSocialMediaStreams
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationSocialMediaStreams

4.2.4   Social media trackbacks

Rhine users were able to share links to Europeana objects via a number of social media sites. This requirement is to show in the Europeana object display how many times an object has been linked to from Twitter and Facebook.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsObjectSocialMediaTrackbacks
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationsObjectSocialMediaTrackbacks

4.2.5   External services integration (via SIWA)

The SIWA framework will enable users to access external services with a single click.

Examples:
- see if a book title is available in a bookshop or library
- navigate to a Wikipedia subject page relating to the object
- search for images in Flickr.

Note: this is a flexible framework which will allow new services to be easily added or removed.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsExternalServicesIntegrationThroughSIWA
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationExternalServicesIntegrationThroughSIWA

4.2.6   Wikipedia-enriched search results

The general idea of this requirement is to provide Europeana users with contextual information when searching for places and people. This feature will combine Europeana search results with contextual information pulled in from Wikipedia via dbpedia, thus providing context for the user and improving Europeana's search engine rankings.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchResultsWikipedia
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationSearchResultsWikipedia

4.3   Facilitation of content re-use

4.3.1   Cite record

This will enable the user to:

- click on a *Cite record* link or button in the Europeana object display, and
- view a new dynamic window displaying a selection of formal citations in different styles for them to copy.

Notes:
- We will initially provide Harvard style citations and subsequently, if possible, introduce citations in other styles

- We will maintain statistics on how many times the feature is used per record.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchRetrieveCiteRecord
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationCiteRecord

4.3.2 Improved search API

A first simple search API, based on the OpenSearch protocol, was released for pilot testing in Rhine. For Danube, we will update and refine this API to reflect:

- the results and feedback from the pilot testing and the development of Europeana search functionalities, and
- the change in the data model from ESE to EDM.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchImprovedAPI
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationSearchImprovedAPI

4.3.3 Embed object (HTML-snippet)

This feature is aimed at bloggers or website creators who:

- want to include individual Europeana objects in their blogs or webpages, but
- do not have the skills to do this using an API.

Europeana will create ready-made HTML-snippets which the website author can copy and paste into their websites.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsSearchRetrieveByHTML
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationEmbedRecord

4.3.4 Semantic markup in object pages

The addition of semantic markup to HTML object pages makes normal HTML result pages machine-readable. This allows specific harvesters to take information from them.

It is more efficient for many search engines if the normal document and the formalised data are located in the same place and shipped along the same channel. This is especially true for search engines or widgets such as the *Like* option in Facebook. **Note**: semantic markup is sometimes advertised as a means of improving SEO.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsEDMSemanticMarkup
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationsEDMSemanticMarkup

4.3.5 Publish Linked Open Data

The goal is to implement the principles of Linked Data to re-publish Europeana data, or a relevant subset of it. Linked Data is one of the rationales for EDM, as it enables the easy connection of one Europeana object to other Europeana objects, or to any other contextual entity.
Note: EDM adoption is not a pre-requisite of Linked Data implementation. ESE can support the publication of certain basic Linked Data.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsEDMLinkedData
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationEDMLinkedData

4.3.6 Content re-use statistics (API, Share/SocialMedia trackbacks, EmbedHTML, CopyPaste)

The ability to reuse Europeana content outside the portal is due to be implemented before the Danube release. This requirement is to introduce tools for collecting re-use statistics, which will be used for reporting and to improve future development planning.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsContentreuseStatistics
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationContentreuseStatistics

4.4 Improved content ingestion

4.4.1 United Ingestion Toolset

This requirement is to streamline the standalone tools we use for content ingestion, enrichment, quality control and reporting into a single workflow. This will reduce the workload for Europeana and its partners and improve the quality of the ingested content.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsContentInTools
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationContentInTools

4.4.2 Content Reporting

This requirement is to introduce reporting tools to improve the quality and collection of statistics on our content ingestion processes, services and products.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsContentInReportingQuality
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationContentInReportingQuality

4.4.3 Customer Service Centre

Ths requirement is to offer a dedicated customer service platform to provide a shared space for our partners, aggregators and content providers. This will enable them to:

- find documentation easily
- refer to FAQs and best practice guidelines

- put questions to the Europeana ingestion team and to each other, and
- view the answers in a single space.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsContentInExtras
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationContentInExtras

4.4.4 Ingestion of contextual content

This requirement is to enable the ingestion of contextual content.

Example: for a painting, information about:

- the painter
- the place of creation
- the subject
- the genre and its place in art history.

This will enable Europeana to present content objects in their context, thus providing the materials needed to build narratives around them. It will also make for richer search and retrieve functions as described earlier in this document.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsEDMContextualDataIngestion
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationContextualDataIngestion

4.4.5 Data Enrichment

This requirement is to enrich Europeana data by means other than straightforward content ingestion. It is particularly important in terms of the search and browse requirements.

Example:
- a user browses facets or maps and finds an object located in Paris
- the Europeana metadata record uses not the string *Paris* but the Geonames code http://sws.geonames.org/2968815/
- the user can:
  - see the word *Paris* in their native language
  - relate the term *Paris* to other places in France **Example**: Ile de France
  - view information on Paris such as geographical coordinates and administrative areas.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsEDMDataEnrichment
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationEDMDataEnrichment

4.4.6 License selection tool

The license selection tool enables the hundreds of Europeana content providers to choose between the several different rights statements allowed in the EDM (from ESE v3.3).

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsContentLicenseSelectionTool

- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationLicenseSelectionTool

4.4.7 Public Domain Calculator

The Public Domain Calculator tool assists the hundreds of Europeana content providers to determine which copyright is in effect on a particular content object, based on the features of the content object and the international regulations.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsContentPublicDomainHelperTool
- Functional specification:
  http://europeanalabs.eu/wiki/DanubeFunctionalTechnicalSpecificationPublicHelperTool

4.4.8 Language Resource Register

The Language Resource Register will help Europeana partners and others to find language resources and information about them. **Examples**: languages covered, license form.

- Requirement specification:
  http://europeanalabs.eu/wiki/SpecificationsDanubeRequirementsMultilingualityLanguageRegister
- Functional specification:
  http://europeanalabs.eu/wiki/SpecificationsRhineTechnicalSpecsMultilingualityLanguageRegister

5 The Europeana Data Model

The Europeana Data Model (EDM) defines the structure of the Europeana data.

The EDM is set out in two separate documents:

1. the EDM specification: this document sets out the formal definitions of the classes and properties that make up the model;

2. the EDM Primer: an informal introduction to the model that uses examples to illustrate its main classes and properties.

This section is designed to present the EDM in a context that will aid understanding of what the model entails. It includes:

- rationale for the shape and form it currently takes
- a brief history of previous EDM iterations
- a brief account of the evaluation of the model to date.

5.1 Rationale

The development of EDM has been guided by three main aims:

- to achieve the integration of metadata from the four main Cultural Heritage domains: libraries, museums, archives and audio-visual archives

- to support rich functionality, notably semantic search, and
- to optimise the use of resources in time.

5.1.1 Cross-domain Data integration

Europeana is a data integration system that can be seen as a living organism. It consists of one central repository and many local sources, where the local sources are the collections held by the content providers that contribute to Europeana.

Europeana is in a state of continuous expansion for the following reasons:

- New data is expected to come from local sources in the form of additions or modifications to data contributed previously
- New local sources are expected to join the Europeana network
- New users are expected to join the community accessing Europeana, either directly via its portal, or indirectly via an intermediary
- New functionality will be required by these users, because the context in which they carry out their activities constantly changes.

Given the particular circumstances, there are two possible data modelling approaches:

- to define a set of cross-domain elements, in the sense of a relatively small, common set of properties capturing features shared by all objects. **Example**: the Dublin Core Element Set.
- to define an ontology, in the sense of a **complete** conceptualisation of the fundamental notions around Cultural Heritage Objects. This would allow Europeana to accommodate providers' data **regardless of the original models**.

The former, cross-domain approach was followed in the Europeana Rhine Programme, of which the principal aim was to set up a basic infrastructure for the collection of millions of objects. This approach has led to the definition of the Europeana Semantic Elements (ESE).

5.1.2 Support rich functionality

The aim of Danube, however, is to offer a far richer functionality, beyond the capacities of ESE. In technical terms, the requirement for EDM is that it contains the expressive machinery to:

- represent complex data objects, resulting from the collection of the rich information of the content providers, and
- connect these descriptions to each other in a meaningful way.

These two aims can only be achieved by the use of a basic ontology acting as integration medium. This is why we chose the ontology approach for Europeana Danube.

The ontological structure of the EDM is a necessary but not sufficient condition for the successful integration of the information collected by Europeana. On the operational level, a further process is necessary; the transformation of the values of the collected metadata in order to harmonise them from the lexical point of view and enhance their quality. This process, *Data Enrichment*, is crucial to the development of Danube. EDM supports data enrichment by defining the classes and properties for the enrichment process. The main burden, however, is clearly the enrichment process itself.

5.1.3   Optimising the use of resources in time

Once the ontology choice was made, we needed to assess how best to achieve the two aims stated above within the resources of the project, while at the same time protecting the investment we are making now.

The only possibility was to minimise the effort expended by building EDM as far as possible on the basis of existing models, and protect the investment with a reliance on standards and current best practices.

An increasing number of cultural institutions are using the web as a knowledge exchange medium, expressing their data and their Knowledge Organization Systems by using URIs, RDF/S, SKOS and OAI-ORE and making them available on the web as Linked Data. It is clear that Europeana can optimise and protect its investment by using Web Architecture and standards in such a way as to exploit the achievements that cultural institutions have already made, not by pushing them in a different direction.

5.1.4   Summary of requirements

In summary, our requirements analysis indicated that the EDM should:

- use a simple ontology to capture all the relevant properties of Cultural Heritage Objects
- offer primitives for collecting complex data from contributors
- integrate all the providers' data
- re-use existing ontology and models, and
- be based on Web architecture, languages and practices.

None of this was clear at the outset. These decisions were made as the result of a process that required several months of collaborative work between the teams, the core experts and all the WP3 participants.

5.2   Brief history

The version of EDM to be released with the Danube specifications is 5.2.

Version 1, released in June 2009, was based on the notion of surrogate and endowed with a rich set of contextualisation properties largely inspired by CIDOC CRM [1]. Its revision by the Office and the WP3 core experts led to version 2, released in July 2009.

[1] Martin Doerr. *The CIDOC conceptual reference model: An ontological approach to semantic interoperability of metadata*. AI Magazine, 24(3):75–92, 2003.

Version 2 was analysed by the experts in EuropeanaConnect, notably the Free University of Amsterdam. As a result, version 3 was produced in early September 2009. Version 3 was closer to the web vision of resources and included OAI-ORE aggregations (and their properties) and SKOS concept.

Version 3 was presented and discussed at the first Europeana Plenary Meeting in The Hague in late September 2009. At that meeting, a session was devoted to EDM, during which comments were made which led to version 4 of EDM.

In version 4, the notion of surrogate was dismissed; the main distinction in the model was that between information and non-information resources.

Until this point, we had been using ESE to structure the Europeana data. However, by this stage, we considered that EDM had matured sufficiently to attain the status of Europeana data model. For this role, it had to be integrated with ESE.

Integration of ESE in EDM led to version 5, released at the end of 2009. With this version, Europeana entered an initial evaluation.

5.3 Initial evaluation of EDM

In the first quarter of 2010, EDM version 5 was evaluated in the context of a series of meetings with experts from the four domains contributing to Europeana: libraries, archives, musuems and audio-visual archives.

A detailed account of these meetings can be found at:
http://europeanalabs.eu/wiki/WP1.

The evaluation led to a minor revision of EDM, which is the present version of EDM. In parallel, a prototyping activity has been launched with the aim of building and testing the basic blocks for managing data in EDM.

6 Future development

This section looks at Europeana development post-Danube.

There are three main considerations for the medium and long-term:

- Resourcing
- Topics for future review
- Other Europeana project outcomes.

6.1 Resourcing

The requirements analysed in these specifications may not all reach the development stage of the Danube Programme. As stated in Section 2, the Business Team will assess expectations and resources before taking the final decision on which requirements will be implemented in Danube.

The functionalities and features specified in this document but excluded from Danube are natural candidates for future developments.

6.2 Topics for future review

Deliverable *D3.3 Initial Technical & Logical Architecture and future work recommendations* sets out a number of topics to be considered for the future. These are:

- Further evolution of the EDM, including FRBR harmonisation
- Linked Open Data Integration and Linking Dbpedia
- Use of DDC as contextualisation resource
- Enable Support for Scholarly Inferencing
- Authentication and authorisation.

Note: The deliverable can be found on https://version1.europeana.eu/group/europeana-collaboratory/documents following the path: Folders » WP3 » Deliverables » D3.3 Initial Technical and Logical Architecture and future work recommendations.

The topics listed above still need consolidation. Each will be specified more fully in Deliverable D3.4, after which the Office will consider them for inclusion in the development process. Those that are included will become formal requirements and go through the functional and technical specification process and on to possible development and implementation.

6.3　Other Europeana project outcomes

We will also be looking at the outcomes of other, ongoing projects in the Europeana group, some of which may take Europeana in completely new directions. **Example**: the multimedia functionalities offered by the ASSETS project. These projects may open up new perspectives that need to be considered when planning the future of Europeana.