Grant Agreement 297292

EUROPEANA INSIDE

Requirements Analysis

Deliverable number	*D2.1*
Dissemination level	*Public*
Delivery date	*August 2012*
Status	*Final 1.0*
Author(s)	*Sam Alloing (KUL), Berry Feith (DEN), Wietske van den Heuvel (DEN), Norbert Kanter (ZET), Monika Lechner (DEN), Jef Malliet (Provincie Limburg), Marco de Niet (DEN), Eva van Passel (IBBT), Marco Streefkerk (DEN)*

This project is funded under the
ICT Policy Support Programme part of the
Competitiveness and Innovation Framework Programme.

Revision History

Revision	Date	Author(s)	Organisation	Description
0.1	2012-09-07	Sam Alloing Berry Feith Wietske van den Heuvel Norbert Kanter Monika Lechner Jef Malliet Marco de Niet Eva van Passel Marco Streefkerk	KUL DEN DEN ZET DEN Provincie Limburg[1] DEN IBBT DEN	First review draft
0.2	2012-09-09	As above	As above	Second review draft with overall copy-editing and layout by Carolien Fokke (CT)
0.3	2012-09-19	As above	As above	Third review draft with changes by Gordon McKenna (CT)
1.0	2012-10-12	As above	As above	Fourth and final review with changes by Berry Feith and Marco Streefkerk (DEN)

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

[1] Not in the project consortium. Subcontractor Catholic University of Leuven (KUL).

Contents

Introduction

Background

The Europeana Connection Kit (ECK) is currently conceived as a set of components, such as Open Source software modules, procedural guidelines, documentation and tutorials which will be designed to make the supply of data to Europeana, and other targets, easier. It will guide its users through the process of providing metadata to Europeana and other targets (e.g. various aggregators). The possible re-ingestion of enriched metadata from Europeana into the content providers' holdings will also be looked at.

Role of the deliverable in the project

The purpose of this task is to contribute to the conceptual model the range of functions which the ECK needs to support in order to automate (or part-automate) the submission of content from collections or content management systems and aggregator repositories into Europeana or elsewhere. The deliverable is an important part of the process of modelling the functional requirements of the ECK.

Approach

Most sections of the deliverable have been based on input from project partners. The first Europeana Inside networking event, which took place in Berlin in July 2012, provided much input for it. In addition, in order to gain more input from all project partners, a general specifications survey was distributed by DEN before this meeting, followed by a partner-specific requirement analysis task assignment afterwards, also distributed by DEN. As such, this deliverable combines the most important stakeholder viewpoints.

Structure of the deliverable

The deliverable starts with a description of the context of use in which the ECK will be deployed, with a focus on the value network and its stakeholders. It then continues with the ECK definition and general assumptions, including assumptions from the main stakeholder groups' point of view. Throughout the deliverable attention will be paid to requirements and assumptions that are specific for certain stakeholders. This is a result from the consultative approach taken to this deliverable as described above. It will illustrate that compromises might be needed in order to reconcile different points of view on what the ECK entails and what it should enable stakeholders to achieve. This may have an impact on the ECK essence and definition, which needs to be open for adjustments if they prove to be required in later stages of WP 2 or perhaps the project as a whole.

The assumptions will be followed by a list of important functional software requirements, broken down according to the main steps in the data provision workflow and again occasionally integrating or highlighting differing opinions from stakeholders. The 'Functional Business Requirements' are an important reminder that Europeana Inside does not only wish to remove technical barriers, but also organisational, financial and legal barriers. The concluding section in this report is formed by the 'Conclusions and Recommendations'.

1 Context of Use

1.1 Europeana Inside within the digital cultural ecosystem for Europe

In order to outline the requirements for Europeana Inside and its ECK, a first step is to look at the context in which the ECK will be deployed. This includes, most notably, the Europeana aggregation landscape and its stakeholders. However it is useful to extend the contextual overview to what can be described as the 'digital cultural ecosystem for Europe', in which Europeana is being increasingly seen as the central portal (see e.g. European Commission, 2011).

The value network around Europeana is complex and layered. It includes many stakeholders in different configurations and interdependencies. The cultural institutions that contribute to Europeana are not only very diverse, but can also follow various paths to become a contributor.

Previous research has already mapped the Europeana aggregator landscape in detail. A report, published in July 2012, in the context of Europeana v2 focused extensively on the value network or partner network in order to outline Europeana's partner strategy (Friberg, Smith & Brunelle, 2012). The report was written from Europeana's point of view but it also takes other stakeholders' requirements into account, and as such, it is highly relevant for the context of Europeana Inside.

This section does not intend to duplicate the efforts of the Europeana v2 report, but it will complement it with value network approaches. In other words, the focus will be on the Europeana value network as a whole rather than Europeana itself. As a result we will look more directly and more in depth at the roles of other stakeholders within the Europeana Inside project and beyond.

From the perspective of Europeana Inside, there are four main parties involved in this value network that contribute to the creation of end user services:

- Content providers;
- Aggregators;
- Collection management software vendors;
- Europeana.

We will discuss these below. This value network description will serve as a background to the requirement analysis of Europeana Inside.

An overview of the main stakeholders, the most important business relations and interdependencies between them is a necessary first step to gain insight into the workflows of the delivery of digital content to Europeana, and to define requirements that need to be taken into account within Europeana Inside. This is equally important for the business requirements that will be outlined further on in this deliverable.

Removing organisational barriers is described on a high level in the Description of Work (DoW) of Europeana Inside. Here it will be detailed in terms of business requirements. This implies an intricate balancing act to meet the requirements of all stakeholders in the value network: the project should bring added value to the stakeholders as much as possible and should avoid weakening any stakeholder's position. Changes in the value network configurations brought about by the project should not only be beneficial to Europeana itself, but also to Europeana's key stakeholders. As the business requirements need to anticipate long-term perspectives beyond the project's scope, it is necessary to also look beyond the project's direct stakeholders.

1.2 Content providers

Europeana currently provides access to collection information from more than 2200 institutions across Europe. There are two main workflows through which an institution can contribute content to Europeana:

- Directly to Europeana;
- Through one or more aggregators.

"*Europeana has approximately 130 direct metadata providers: 1/3 of it is single institution providers and 2/3 is aggregators*" (Friberg, Smith & Brunelle, 2012, p16).

Although Europeana is encouraging single content providers to deliver their content first to an aggregator, they recognize from certain content providers' point of view that contributing to Europeana without aggregators as intermediaries may be a requirement: "*In cases when a new partner wishes to represent content via Europeana but cannot be redirected to one of the four aggregators roles (…)*[2]*, Europeana will take the content directly*" (Friberg, Smith & Brunelle, 2012, p48).

There are a few reasons why there are single content providers. One of them is that Europeana would like to have prestigious cultural heritage institutions on board as an incentive for other institutions to join. The Rijksmuseum Amsterdam (RMA) for instance is currently one of the single content providers in Europeana for that reason. Europeana has provided customised services to the RMA to get their metadata on board.

Another reason why there are single content providers is that there may not be suitable aggregators through which an institution can push its metadata to Europeana. This can be caused by various circumstances:

- Lack of an organisational infrastructure for aggregating in a country;
- Lack of a matching thematic profile of existing aggregators;
- Lack of appropriate technology (e.g. harvesting) at the institution to participate in an aggregator.

These are all relevant scenarios for Europeana Inside to take on board.

[2] In this quote, the four roles are national initiatives, national aggregators, domain aggregators or project aggregators. The strengths and weaknesses of all these approaches, not only for Europeana but also from the aggregators' and content providers' point of view, are described in detail in the report (Friberg, Smith & Brunelle, 2012, pp21-22).

The opposite happens as well. A content provider, able and willing to channel its metadata to Europeana (or other information services) through an aggregator, may decide to participate in more than one aggregator, as different aggregators may provide different values and services. A content provider may distribute different digital collections across different thematic aggregators. Or they may participate in a national aggregator with its entire digital collection and also with sub collections in an international thematic or domain aggregator. This may result in the duplication of records for the same thing in Europeana. This will be solved if all content can be recognised by a unique identifier, but this has not yet been achieved. Content providers obviously need to make clear decisions on and agreements with aggregators about which content will be contributed to Europeana and through which channel.

1.3 Aggregators

An aggregator collects metadata about heritage documents or objects from several content providers, assembles it in a uniform format and makes them available for use and reuse in other contexts, according to agreements with the content providers.

In the context of Europeana the aggregators are used as a supporting organisation between Europeana and content providers. They can offer:

- Technical support and advice to content providers;
- Technical services such as conversion of metadata or transfer facilities according to various protocols;
- Internet publication of the metadata and digital objects from their content providers as a service;
- Long-term digital archival functions to their content providers;
- Functionality for enriching and harmonizing the metadata provided by their content providers;
- Administrator operations;
- Training.

Europeana can build a long term relationship with aggregators. This makes it easier to communicate (for example about the guidelines of Europeana).

It is obvious that the role of aggregators is very important in Europeana's ecosystem and the intermediate aggregator track[3] is Europeana's most important contribution method. Europeana prefers this for reasons of efficiency and cost effectiveness, and therefore its strategy includes encouraging individual content providers to provide content via suitable aggregators.[4]

The varied aggregator landscape described in detail in the *Europeana Partner Strategy and Development Plan* is still expanding, and recent EU policy initiatives stress the importance of aggregators as well (e.g. European Commission, 2011).

[3] The term 'intermediate' is used in this context to recognise that Europeana can in fact be seen as an aggregator in itself. When the term aggregator is used in the rest of this report, it will refer to the definition that is customary in Europeana related terminology, i.e. aggregators for Europeana, not including Europeana itself.

[4] See the *Europeana Content Strategy, Partner Strategy and Development Plan and Collections and Data Analysis Strategy and Plan*.

Contributing content to Europeana is generally a major service that aggregators provide, but in most cases it is not the only one. Aggregators also provide other services, e.g. public-facing or business to business, and can be used in other situations. Some aggregators have a public front-end, such as a thematic portal, with its own value propositions. Others have no public interface and were installed only to collect and standardise metadata for delivery to Europeana. However this has not fostered a situation where all cultural institutions in Europe have easy access to a suitable aggregator. Moreover, the long-term sustainability of some aggregators, most notably project-based ones, is a major concern, as voiced by the Europeana Financial Sustainability working group.[5]

Different types of aggregators

The table below shows the types of aggregators distinguished within the Europeana aggregation infrastructure.

Domain	**Cross**	Aggregate metadata across domains e.g. Erfgoedplus.be
	Single	Metadata from a single domain at regional, national or international levels e.g. Apenet Project
	Thematic	Collections of Jewish culture project Judaica
Geographic	**Regional or national**	Metadata at a regional and national level can be either single domain, e.g. national library, or cross domain, e.g. Culture Grid, run by Collections Trust in the UK
	Pan-European	Represent a specific segment or sector of cultural heritage by aggregating metadata on a European level, e.g. TheEuropeanLibrary.org
	National Aggregation Initiatives	Organisation(s) appointed by their ministry to take on the role as aggregator in the country, e.g. these initiatives can be either cross- or domain-specific organisations. Example of cross domain is Hispana.mcu.es
Business	**Projects**	Temporary funding to deliver content for a fixed period of time, e.g. http://www.athenaeurope.org/
	Infrastructural	Initiative with (semi-)permanent public funding as part of (national) infrastructure. No direct user service, e.g. http://digitalecollectie.nl/
	Sustainable	Aggregators that generate revenues (commercial or public) to cover the exploitation costs, e.g. http://www.scran.ac.uk/

Extension to: http://pro.europeana.eu/web/guest/aggregators-and-providers

[5] http://pro.europeana.eu/documents/900548/1114203/Financial+Sustainability+Working+Group+Report. See also the strengths and weaknesses of aggregator types (see next footnote).

Added value of aggregators

Technical value

Aggregators may handle more extensive and complex metadata models than the Europeana Data Model. This has the advantage that they can build specific portal sites that can give better and more comprehensive search results and information, in their own specific context. The rights on the metadata may be different from those in Europeana, which means content providers may be less reserved than delivering metadata to Europeana.

Aggregators can combine information from different sources like Europeana does, and therefore they can also enrich metadata, and possibly send those enrichments to Europeana. They can be partners with Europeana to give richer records, and if you can enrich sooner you can get better results. Content providers and technical partners would like to study more concrete examples of data enrichment by aggregators: how valuable this is, how is it realised, in what form and how is it maintained?

It is an advantage if the enrichment already starts at the aggregator (and is included in the ECK). In many cases enrichment of metadata is easier and more reliable if it can be supported by the authors of the source metadata (the collection holders). Also for this purpose, the aggregators are better placed to lead the process, as they are closer to the content providers. Summarized: enrichment is enhanced when aggregators participate in the process.

Because an aggregator can develop, depending on the type of aggregator, a long term partnership with Europeana, and because an aggregator has the opportunity to work with various types of source metadata, mapping and harmonisation will be easier and more consistently done by aggregators than by individual content providers. An aggregator could help Europeana's long term strategy to provide PIDs and a resolver infrastructure for PIDs.

Content providers need less technical knowhow and infrastructure. An aggregator can take the role of technical partner to support the content provider not only with mapping the local metadata to the aggregators data model or the Europeana Data Model, but they can also have technical infrastructure to deliver metadata to Europeana, store objects like thumbnails or have infrastructure for enrichment of records.

Business value

Aggregators deliver business value by:

- Being closer to the content provider than Europeana, which means that they can give quicker and more specific support;
- Supporting the network of joined institutions to spread their metadata to a broader audience and create a bigger involvement in Europeana;
- Promoting and implementing metadata standards on a broader scale. Metadata that will be used in a global repository needs to have more metadata and needs more information more explicitly, like the country of origin, than if the information stays local in the CMS[6].

[6] Note that the abbreviation CMS in this document always stands for 'collection management system' and not 'content management system'.

- Supporting the organisations by bringing together the experience of other participating organisations (depending on the scope of the aggregator);
- Representing their content providers at Europeana meetings on technical or organizational issues.

Requirements for aggregators

An aggregator has the same added value as the systems used by Europeana but on a smaller scale:

- Mapping tools for metadata (from the local metadata fields to the data model used by the aggregator);
- Metadata quality control;
- Delivery and exchange infrastructure;
- Publishing infrastructure;
- Enrichment infrastructure.

Aggregators usually already have their own infrastructure in place, their own workflow set-out. Using a generic ECK has several benefits for them:

- The barrier for institutions to provide content to the aggregator will be lower when they have a Europeana Inside CMS;
- The support of the software will be shared by a larger user community of aggregators;
- The knowledge sharing will be more effective.

All in all, it is important to make a distinction between content providers and aggregators in the use of the ECK:

Content providers

- Use the ECK mainly for preparing metadata from their own collection registration systems for automatic processing by one or more aggregators;
- Are responsible for the content of a metadata record;
- Are experts on the content of the digital object;
- Have the final authority regarding legal issues (e.g. copyright, licensing) about the source metadata.

Aggregators

- Will mainly use the ECK as a platform for automatic validity checks and adaptation before forwarding the metadata of a single institution or an aggregation of metadata from various institutions to Europeana;
- Have the technical expertise regarding metadata formats and other technical information;
- Are also more knowledgeable about organisational issues, such as IPR or business models for digitisation and digital heritage.

1.4 Technology providers

It is important to note that Europeana does not directly list Collection Management System vendors as stakeholders. They might be inferred in the category technology organisations, whose stake in Europeana includes technology developments and knowledge exchange. Within Europeana Inside CMS vendors take on a more prominent role than is usually the case in descriptions of the Europeana landscape and they will have to be included separately in the value network. Technology is often implicitly deemed to be a part of the role of content providers and aggregators, as is clear by the explicit mention of developments in technology in the description of e.g. the cultural institutions stakeholder group (Friberg, Smith & Brunelle, 2012, p. 18).

Europeana Inside and its ECK can potentially strengthen the link from content providers to Europeana through the CMS vendors, who will hence get a more prominent role in Europeana's value network. An important aim of this, as described in the project's Description of Work (DoW), is to reach the vendors' broader user-base, including institutions that lack in-house technological expertise to correctly provide metadata to Europeana. For instance, the ECK will facilitate mapping to the most up to date Europeana metadata models in a manner directly linked to the institutions' software environments. From that perspective, there are two main scenarios' that should be taken into account:

- From a technical standpoint, the aggregator can be considered technical (and organisational) stakeholder. It is the stakeholder that supports the content provider to contribute to Europeana. If the content provider wants the CMS vendor can be involved by supporting them to deliver the content to the aggregator. The vendors know the local situation; they can for example export the metadata from the CMS to a standard metadata format.
- The vendor and the ECK can take over some of the roles of an aggregator, for instance for content providers that are not yet part of that aggregation landscape.

However it is useful to stress that the project should not lose sight of providing added value for aggregators as an important stakeholder category either. Technical facilitation of the mapping process will also alleviate the workload of many aggregators and increase efficiency to enable them to focus on other essential tasks or aggregate content from a higher number of providers than was previously possible. The ECK can bring added value both to aggregator scenarios and direct contributions scenarios, but striking the right balance will be important.

1.5 Europeana as a partner in the value network

An important challenge for Europeana Inside will be to match the organisational requirements of Europeana, most notably being able to manage incoming metadata from either aggregators or individual institutions. The project wishes to reduce or remove organisational, technical, legal and financial barriers to participation. In order to reduce organisational barriers to the largest possible extent, the project needs to look at different existing institutional scenarios and foresee the largest feasible number of different road maps to match as many of the scenarios' requirements as possible.

The ECK should be applicable both in aggregator scenarios and in direct contributions scenarios. However, it is also important that opening up these additional opportunities does not entail a weakening of the position of existing aggregators, who have proven to be very

valuable in assisting institutions with content provision and in streamlining Europeana's organisational work processes. This illustrates the delicate value network balances in place and the need for Europeana Inside to address these issues as well as the important technical challenges. This largely non-technical contextual introduction to the requirement analysis, looking at the digital cultural ecosystem surrounding and contributing to Europeana, is useful to address potential imbalances in the value network that might arise.

A useful way to map the value network as a whole is the Europeana partner strategy plan. Europeana takes a business modelling approach in its strategic plan and distinguishes its partners from its stakeholders, whereby the main partner group includes content providers and aggregators, and the main stakeholders can be summarised as users, policy makers and market players. The 'business modelling canvas' approach based on Osterwalder and Pigneur (see Osterwalder & Pigneur, 2009; Verwayen & Limonard, 2010) departs from a single organisation's position more than the value network approach taken in this report, so it is highly suited for Europeana's purposes and is used again in the partner strategy plan (Friberg, Smith & Brunelle, 2012, p11).

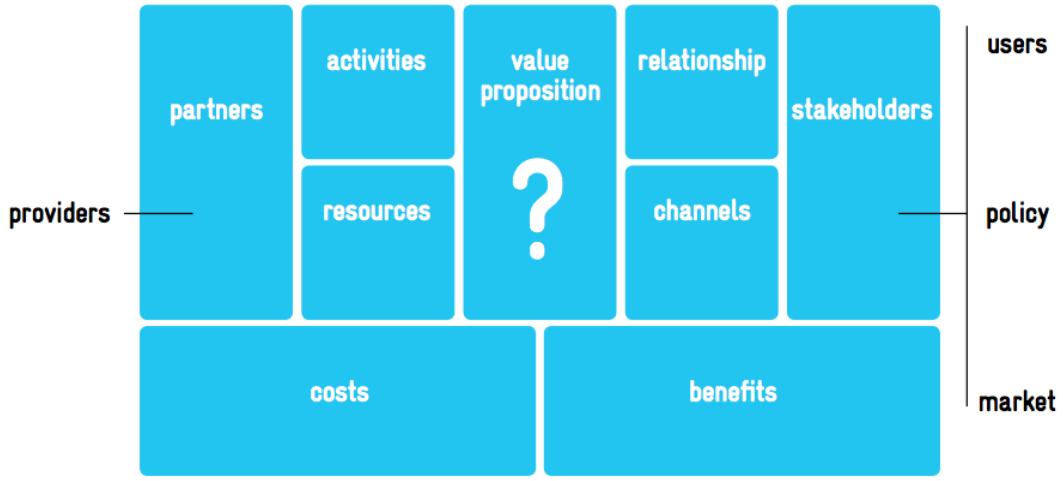

Figure 1 – Business Model Canvas developed by Osterwalder and Pigneur (Europeana, 2011, p6)

It is immediately clear that this is a simplification of a complex landscape. Europeana has since expanded its own stakeholder list greatly to a list of 19 different stakeholder types, which includes not only current but also potential stakeholders:

- European Commission;
- Associations in public sector;
- Ministries (of Culture, Education, Tourism, ICT, Infrastructure);
- Cultural institutions (GLAMs);
- National initiatives;
- Single domains (TEL);
- Cultural portals;
- Project partners;
- Social media and community organisations;
- ICT and content seeking companies (telecom, IT, web based organisations);
- Tourism agencies;

- Education and learning sector (universities, science, schools);
- Publishers and news agencies;
- Broadcasters;
- Creative industries;
- Sponsors;
- Private contributors;
- Technology organisations and advocacy stakeholders.

(Friberg, Smith & Brunelle, 2012, pp. 18-19)

We will not go into this list in detail, and we refer to the source report for the description of the nature of the stake in Europeana of all these actors, but it is clear that the list is a very useful source to map the value network.

1.6 The value network

Based on a combination of the stakeholder list provided by Europeana and Europeana Inside's key stakeholders, a value network representation of the Europeana landscape can be devised. It represents the most important stakeholder categories and including Europeana as one of a number of distribution channels.

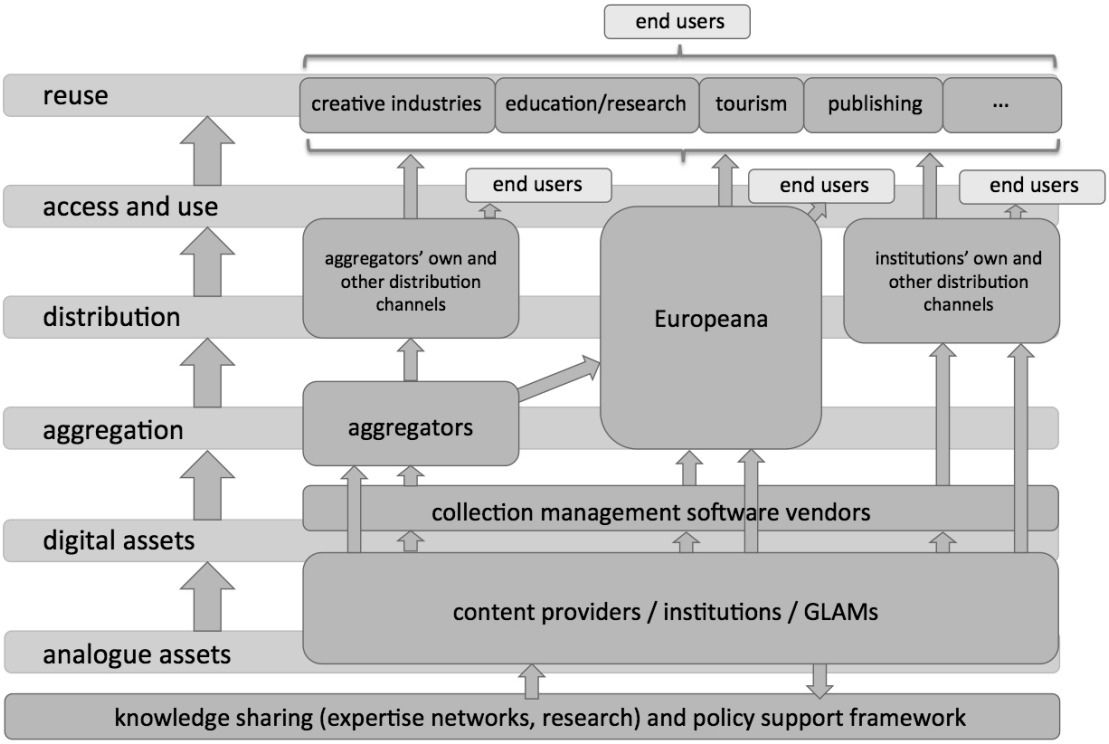

Figure 2 – Europeana Value Network for Europeana Inside[7]

[7] This value network is partly based on the research project Archipel in Flanders, Belgium, where a strategic road map was devised that includes the Archipel value network (Van Passel, 2011b). It has of course been adapted thoroughly to reflect the Europeana context. It for example includes a reuse phase, an important strategic goal for Europeana, and also discounts the phase of preservation, which was important within the Archipel content but is not relevant for Europeana Inside. The value network approach is based on the business modelling methodology as applied within Archipel (Faber et al., 2003; Ballon, 2007; Ballon, 2009 as cited in Van Passel, 2011a).

This value network, while still a simplification of a complex landscape (e.g. by grouping all aggregator categories in one stakeholder group), sufficiently illustrates the most important business relations from each stakeholder's point of view and the different potential scenarios for content providers to contribute to Europeana. It is a visualisation of the different paths content providers can take to contribute to Europeana. It must be noted that the content provider, aggregator and CMS Vendor categories all include partners from within Europeana Inside as well as other providers, aggregators and vendors.

An important remark with regards to this visual representation is that it currently only reflects the unidirectional flow of content from content providers to Europeana, other aggregators and beyond, including to stakeholders that are expected to reuse the metadata.

Europeana Inside also foresees a 'return workflow' (cf. infra) that would allow ingestion of enriched metadata for collection providers' digital assets. This can be considered an important added value and incentive for content providers to participate, as is described in the project's DoW, and can be highly useful in the business model development of institutions (cf. infra). However, as will be detailed later in this document, it will not necessarily be the case that the return workflow will follow the same path as the uploading workflow.

The return of enriched metadata is still in an experimental stage, and the workflow has not yet fully crystallized. As such, for this figure's clarity, it was decided not to include the bi-directional flow of metadata at the moment, which does not imply it will not be taken into account.

Aside from the necessary balance between scenarios with and without aggregators, both represented in the figure above, a balance between interests of the CMS vendors and content providers is also expected to be an important point of attention throughout Europeana Inside. CMS vendors offer a commercial product and service and monetising extra functionalities related to the ECK may seem logical from their point of view.

From a content providers' point of view, but also from Europeana's point of view, facilitating contribution is key, which means that extra financial barriers will not be desirable if they are to reach their (non-economic) business goal of increased access to cultural content. This can be expected to be an area of tension that needs to be sufficiently taken into account to maximise uptake if each stakeholder is to find added value from Europeana Inside and the ECK.

Finally, it should be noted that the aggregators' and institutions' own and other distribution channels, while they may not be the main focus of Europeana Inside, are important to consider within the value network, most notably in terms of the desired metadata output of the ECK. It is clear that the ECK will meet more of the aggregators' and providers' business goals, and will hence remove more barriers to participation, if it facilitates not only provision to Europeana but also to other distribution initiatives that institutions and aggregators undertake or wish to undertake in the future.

2 The Europeana Connection Kit: definition, assumptions and quality expectations

2.1 Definition of the Europeana Connection Kit

In the Europeana Inside project proposal, the ECK is envisaged as "*a suite of open-source software tools which will automate, or part-automate (depending on the complexity of the sources data) the flow of content into Europeana*".[8] So the ECK can be defined as a set of tools, including software, procedural guidelines, tutorials, documentation and other components which will help all stakeholders to streamline the supply of metadata to Europeana.

The nature of the individual components has not yet been defined precisely. Sections 2 and 3 provide a first step towards precise definitions by presenting requirements collected from all project partners, content providers and technology partners alike. As this is a user-driven project, the voices of all partners need to be heard at this early stage of the project to come up with a toolset that adds value for all partners and other parties outside the project.

The first step in the requirements analysis to determine the specifications of the ECK was to send out a questionnaire to all Europeana Inside partners to collect information about their expectations and requirements of the ECK. This section provides an overview of the general assumptions about the ECK which were collected by the questionnaire[9]. The next section will address the individual components and functionalities, as seen by the project partners.

2.2 Users of the ECK and their roles

In the previous section, we discussed the main types of users (stakeholders) in the Europeana Inside Value network:

- Content providing institutions (also referred to as 'content providers' throughout the document);
- Aggregators/service providers;
- Software developers/CMS vendors(also referred to as 'technical partners');
- Europeana.

In this section we will focus on them in relation to the ECK. Each type of user was asked for their ideas and expectations, including their own roles in relation to the ECK. The most prominent roles and responsibilities that mentioned were:

- Content providing institutions:
    - Select a route to Europeana (directly or via one or more aggregators);
    - Adhere to standards;
    - Resolve IPR and other legal issues;
    - Quality assessment of metadata and conversion quality;
    - Make an effort to understand the processes.

[8] From Europeana Inside's Project Proposal Part B, p. 9 (Europeana Inside, 2012).

[9] Requirement Analysis Task Assignment (DEN, 2012).

- Aggregators/service providers:
    - Service the ECK;
    - Support their content providers
    - Main role in testing and monitoring;
    - Suggest improvements, new requirements.
- Software developers/CMS vendors:
    - Assess the connectivity of a CMS;
    - Support proper XML exportation as a minimal requirement in connectivity;
    - Enhance user-friendliness and transparency;
    - Acknowledge the collection manager to be in control of the metadata;
    - Improve the application of standards.
- Europeana:
    - Provide proper feedback to the content providers;
    - Timely ingestion processes;
    - Implement changes to data model with care;
    - Maintain the target data model part of the ECK (system and documentation).
- ECK itself:
    - Must streamline the processes;
    - Provide support for non-technical partners;
    - Provide simple and transparent procedures, applications, guidelines;
    - Must always be up to date.

2.3 How content providing institutions imagine the ECK

Streamlining metadata to and from Europeana

The main purpose of the ECK is to streamline the flow of metadata from content providers to Europeana and back. It plays an important role in making this process easier for content providers. It supports the exchanging of metadata between the CMS and itself and between it and Europeana (and vice versa).

Distribute to platforms other than Europeana (e.g. aggregators)

The ECK must be able to push and pull metadata to and from dissemination platforms other than Europeana. Strictly speaking it is a 'Europeana' Connection Kit, but preferably metadata is submitted to Europeana through aggregators, and these usually have other requirements and use other methods and data models, supporting their own business models.

Flexible and versatile

As a 'Connection' Kit, it will have to set basic requirements concerning the connectivity of sources (input) as well as targets (output), and the way in which they are used in a specific collection context. For example how strictly the actual field content corresponds to what is in the specifications (data dictionary). It should be possible to support new standards as input and output and map these accordingly.

Easy to use

The ECK should be a software tool that works without the user realising that he is using it. It should work as much in the background as possible. It does not create new barriers, but eliminates barriers. Each part of the ECK should be intuitive and self-explanatory. However, good user manuals and training materials are needed, both on technical implementation of the tool and on the use of the tool.

Incorporating existing tools

The ECK should include and improve as many relevant existing tools with similar functionalities that are available according to existing best practices, in order to avoid duplicating efforts.

Integrated and an external application

The ECK can be an integrated part of CMS software, but it should also be possible for an institution to use it if it does not have a system that has integrated the ECK. It can still deliver metadata that can be imported in the ECK. This is to keep the barrier as low as possible.

Always up to date

It is only usable if it is updated according to the latest Europeana requirements at all times. Users should be properly and promptly informed of any changes in the ECK.

Version tracking

Versions of different updates must be clearly identified as such. Any process must be designed so it can be repeated (and modified if needed) at a later stage, without confusing old and new metadata. An automatic verification of the latest version of an application should be available when the user runs it.

Open source to prevent vendor lock in

The ECK should not allow lock-in to a vendor system and thus be open source, but someone must be responsible for the maintenance of critical parts.

Documentation

The system must be documented in a user-friendly manner, and documentation must be properly maintained along with the system - which is also a requirement for open source software. The documentation in itself should also be user-friendly and understandable to non-technical people.

Sustainability

The design and development of the ECK must also provide for a sustainable future. After the end of the project, it must constantly be kept up to date.

No or little extra costs for content providers for installation, implementation, use and maintenance

- Collection managers should be able to use the ECK without significant cost increases for their CMS;
- Most development must be done for the common ECK modules, required adaptations in the CMS must be kept minimal;
- Some modification will be required to existing CMSs. These should be limited to general good practice, in view of connectivity, e.g. proper XML export of the metadata from the CMSs belongs to the minimal requirements of a good CMS, and no extra cost should be charged to the collection managers for necessary adaptations;
- Europeana-specific requirements should be solved and maintained in the ECK, not in the CMS.

It is assumed that after a connection has been established between the ECK (tools/framework/ components) and the content provider's software system, it will guide the user (i.e. the content provider) through the process of:

- The delivery of standard descriptive metadata to a target (e.g. an aggregator and Europeana) by a content holding institution;
- The possible ingestion of enriched metadata from the target by a content holding institution.

As an overall conclusion the general view of content providers on the ECK is that it should make it simple for them, especially if they have non-conformant CMSs, to conform to the standards and protocols required by Europeana.[10]

2.4 How technical partners see the ECK

Core assumptions

The core assumption about the ECK is that it should make the process as simple and as fast as possible. It should enhance transparency, so that a content provider always knows the current status of the processing of his metadata. The quality of the content should be key, not only the quantity. Aggregators should have a role in this quality control. An issue that should be clear to the content providers from a technical perspective is the improvement of the speed at all stages. The ECK will only simplify some of the steps in the delivery process to Europeana. Much of the aggregation chain will stay the same.

[10] From Europeana Inside's Description of Work (Europeana Inside, 2012), p. 138.

Using existing software, CMS and infrastructures

When developing the ECK, the existing software, CMS and infrastructures used by the content providers should be used as a basis. Also supporting browser based tools should be provided so that they can operate in parallel with the existing Collection Management Systems. This way, groups of functionalities such as enriched metadata feedback from the aggregator could be smoothly assimilated in the existing workflow of the curators' tasks and activities.

Open source

The common software components should be made available under an appropriate open source license. However, creating no extra cost for content providers might not be feasible. While the ECK will not have a license fee, the service for a single content provider might be charged for the service of a vendor to arrange the connection and mapping to the ECK.

ECK Documentation

ECK documentation should be clearly partitioned along several use cases to be clearly understandable to different types of users. Documentation separated in this manner will not force people to go through documentation that was not intended for them (so it is easy to separate e.g. software integration guidelines from content provider guidelines).

Efficiency

Resource consumption should scale with the number of tools and components of the ECK being used (so unused modules should have minimal or no resource footprint).

Universal tool

Technical partners foresee problems if the ECK will be a universal tool to exchange metadata with any portal. According to them, this opens up a whole new world of interfaces, standards and data formats. It should be a part of the definition that the implementation supports various standards and the processing should be both automatic and manual. The software should be open to new standards and new systems. The project should also regard the limits of Europeana Inside and focus on developing a tool for Europeana.

ECK label

Another assumption from the CMS vendor side is that there should be a label like "Europeana ready" or "ECK validated" for their products. This label should be awarded after some kind of formal procedure.

The role of aggregators

The technical partners do not only have assumptions on the ECK in a strict sense, but also on the role of aggregators in the value network. Most software vendors do not see aggregators as active users of the ECK, since the content providers themselves will use the ECK to send metadata to the aggregators or Europeana. For them, aggregators will only pass on the metadata to Europeana. Their role would be mostly on the administrative side (registering new content providers, forwarding or hosting metadata).

Technical partners stress the need to be clearer about the role of the aggregators. In their view, aggregators are the perfect contact point for any administrative task; registering for Europeana, validation of the content source (is it a museum or a private person?) and communication, but they should not need to manipulate the metadata.

ECK outside CMS

Most, but not all, technical partners prefer the ECK to be outside the CMS. The ECK should be like "*a library/web service outside of the CMS that the CMS can use. This library/web service should be agnostic in terms of technical implementation. Preferably, the ECK library/web service should then be hosted centrally, so no local installations are required at the content provider sites*."[11] Further specified: "The ECK should consist of two parts:

- *ECK local* which is the connector to the CMS and *ECK global*, which is a web service that offers validation and transformation[12].

If aggregators are actively using the ECK, they should be regarded as content providers. This would also be the case, when aggregators enrich the metadata they receive from the content providers[13].

2.5 Additional assumptions

Some assumptions brought to light by the questionnaire cannot simply be ascribed to content providing institutions or technical partners, but refer to compromises between technical partners as service providers and the institutions they provide their services to, or to points of view from aggregators as service providers. As is the case throughout this section, some assumptions seem contradictory in nature.

There are differing views on preview functionalities, which is not simply caused by different views by content providers versus technical partners. Some state that the preview feature should be a generic implementation, more like a diagnostic view, to enable the content provider to verify that the mapping that is in place is correct. Other partners state that users should be able to preview how their information will appear in the target system (e.g. in Europeana). This may be hard to implement, in particular because Europeana may change how it uses and presents the metadata at any time.

The ECK should lower financial and technical barriers specifically, but in fact all barriers in general to transferring information to aggregators and digital libraries (e.g. Europeana). Some service providers refer to this idea from the project proposal to stress that the ECK must not be an extra to a CMS, but should instead be part of the standard package of a good, modern CMS at no extra cost. There can be extra costs for additional services, but one of the aims of the ECK is also to reduce the need for such services.

[11] by Adlib

[12] by Zetcom. See Appendix 1 - Figure 3

[13] by Zetcom. See Appendix 1 - Figure 4

This touches on the question of sustainability and embedding of the various functional components of the ECK. On the one hand taking responsibility for (parts of) the ECK will bring financial costs for service, maintenance and upgrades. On the other hand it will provide the responsible party a strategic position that might bring additional value, social and economical. The Europeana value network is still in its infant stages and all stakeholders are uncertain how much their position will hold or if they need to acquire new roles. Looking from the perspective of the ECK, an initial list of what might be called business requirements in the project phase as well as afterwards for the ECK are:

- Building the ECK: must be complete, no missing pieces;
- Commitment to maintain essential parts, e.g. updates of EDM;
- Commitment to fixing bugs after project;
- Expansion after the project: other scenarios, other domains, other data models, other requirements;
- Maintaining the server and internet domain.

From the composition of the Europeana Inside BPN it is safe to say that we expect that the functionality of the ECK will be split between a local, tailor made part, for which responsibility will be to be taken by the content provider, CMS vendor or service provider, and a global, generic part that will be the responsibility of an aggregator or Europeana itself. It seems in the public interest of the heritage sector to make the generic part as large as possible although it must be acknowledged that at this stage individual content provider often feels the need to put their own interest over what might be called common interest.

This section on additional assumptions also brings to light some minor disagreements that will be detailed further in the next section are:

- The implementation of the ECK should be as simple as possible, making it both easy to maintain and to implement it in CMSs;
- For the uploading and updating of metadata, there should not be human involvement other than perhaps the selection process;
- The project as a whole should aim to get as much content as possible or desired into Europeana;
- The ECK should support different languages, at least the ones involved in the project;
- It should also support different scenarios of exchange and re-ingestion[14];
- The project should aim to make the ECK a standard for Aggregators (for example Culture.fr or Joconde.fr);
- Content providers need to be aware of the workflow timing, i.e. when their content will be available in Europeana;
- Documentation should not be more extensive than required.

[14] Re-ingestion: the transferral and re-integration of, ideally enriched, data back into the content provider's system or database.

The ECK should:

- Be developed using open source methods and multilingual documentation should be available;
- Be implementable with any CMS and is not restricted to the transfer of information to Europeana;
- Connect to the existing CMS;
- Take into account the processes currently in use to provide content to Europeana as these have worked well for existing providers who would not want to 'reinvent the wheel'.

Information should be stored and transferred in appropriate standard formats. This means that the formats are publicly available and are recognised by the community, e.g. LIDO, ESE, EDM, and EAD. There is a set of methods for metadata transfer available (e.g. OAI-PMH) and transfer should be of metadata and not of assets (e.g. image files).

Regarding the transfer of metadata, the issue of control is also important. Until the automated procedure is failure proof, the aggregator and/or Europeana should be in control of the ingestion of metadata from content providers. Content providers should be in control of the re-ingestion of enriched metadata from aggregators and Europeana. The method and formats used to transfer information has to be agreed upon by content providers and aggregators as well as Europeana and vice versa when using the ECK.

2.6 Concluding remarks

Most expectations regarding the functionality and contexts of use of the ECK are shared by all partners. The ECK should make the process of metadata exchange easier and user friendly for content providers. Technical and content partners agree on using the existing software and systems of the content institutions as a basis for the development of the ECK.
There is no agreement between partners as to whether the ECK should be implemented in the CMS or should work as an independent web–based tool. Technical partners indicated it might not be possible to create a tool that can also be used outside the context of Europeana, while content partners feel this should be an important feature.

Another issue on which there is no clear agreement is whether or not the use of the ECK can result in extra costs for the content providing institutions. A possible compromise can be summarised as "there should be no extra costs for content providers, except for additional services" (cf. previous section), but it can be expected that opinions might then differ on what constitutes an additional service.

Another contradiction is related to the preview functionality it should offer, i.e. while preview in general is considered useful, there is no clear agreement on whether or not an "as in Europeana" preview functionality is the best option.

Finally, it is useful to stress that whether or not the ECK will also need to store and hold data (rather than only act as bidirectional channel to pass it on) has not really been explicitly questioned in this section. This needs to be kept in mind for an efficient workflow process.

3 Functional software requirements for the ECK

In order to guarantee an improved workflow for all parties involved in the of exchange cultural content with Europeana there are some general assumptions about this workflow and the functional software requirements of the ECK. A tentative workflow of several steps has been designed to describe the functional requirements for each step.

1. Manage;
2. Select;
3. Prepare;
4. Validate;
5. Supply;
6. Accept;
7. Return.

However these steps need not necessarily be valid for every user scenario. Some steps might be necessary in one scenario, while they are not needed in another. A linear approach is chosen with a maximum number of possibly necessary steps in order to tackle all possibly necessary functional software requirements for the ECK.

In the basic use case with the content provider as start of the workflow and Europeana or another aggregator as end point, the scope of the ECK might be limited to just prepare, validate and supply. Manage and select will be covered by the CMS; supply by Europeana or the aggregator. However in the use case of Europeana providing enriched content back to collection managers the workflow is reversed. In that case aggregators are confronted with requirements to select metadata to send back and the original content providers need functionality to accept the enriched content back into their CMS. This possibly brings an extra range of requirements to the scope of the ECK that need to be analysed. However, we do need use cases from content providers and aggregators to establish requirements.

There are some assumptions about the functional software requirements which unlike most other assumptions cannot be easily assigned to one workflow step in particular. They are given here before each tentative step is described in detail below:

- *Automated workflow:* Create an automated workflow within the ECK. Manual processes and the use of other software should be avoided as much as possible. Also when metadata is being processed, e.g. pushed or pulled by aggregator, it should be possible to logout and check back later to see status.
- *Granularity:* Content providers want to control the metadata that is sent to Europeana i.e. which information to include and exclude.
- *Simple user interface:* The ECK should be simple to use, for example when customizing or reusing metadata crosswalks.
- *Speed of the workflow:* The ECK should not significantly slow down the infrastructure when exporting, transforming or uploading content.
- *Trace workflow steps:* Log files of performed actions on records (like data export) should be available, either directly in the CMS (if the ECK is implemented) or in the ECK. The ability of users to keep notes and document the workflow would be useful. These notes can be used next time an update is performed (there can be a long period between updates).

- *Space efficiency:* The ECK should not require much additional space in the content provider's current technical infrastructure.
- *User roles:* The ECK metadata processing and publishing workflow should define clear roles and responsibilities, for example Administrator, Annotator, Annotator & Publisher, Data Viewer, No role (for rights not yet assigned).
- *Support persistent identifiers*: The creation, management and use of persistent identifiers should be sustained or made possible by the ECK.

3.1 Step 1: Manage

Strictly speaking, the data management itself is handled by the collection holders and thus content providers themselves since it primarily concerns tasks that have to be performed by the CMS. For the purpose of ingestion into Europeana it is necessary to make the metadata available outside the CMS. Therefore steps need to be taken (on institutional and thus CMS level) to transform the internal metadata into externally usable metadata. While CMS software vendors assume that this is outside the responsibility of the ECK, content providers point out that "*the* [CMS software] *tools were rarely or never developed for the purpose of data export and aggregation by Europeana or any other third party.*"[15] Thus content providers need to know what the implications of the transformation, from internally used metadata to externally used metadata are, and be in full control of this transformation.[16]

3.1.1 Moving metadata from a local to a global level

Moving metadata from a local level to a global one requires collection holders to apply widely accepted guidelines and standards. It will also require that collection holders have a better grasp of the structure of their information and how it is modelled in their database (CMS). This transformation touches the following aspects:

Use of controlled vocabulary:

Since the collection metadata is usually aimed at internal use it carries a certain implicit context which will be lost when the metadata is made available in a wider context. Therefore such implicit assumptions will have to be made explicit. The context of the metadata is usually also expressed through specialised language which needs to be made understandable to readers outside of the original context.

When describing assets (digital objects that either represent a physical object or are digital born) the collection owner might or might not use standardised controlled vocabulary[17]. The use of controlled vocabulary is advisable when making the collection available in a broader context, thus outside the CMS, because the semantic linking with other collections or to other thesauri can, hypothetically, be standardised and applied more easily to all collections that use the same thesaurus.

A point for further discussion is whether the cleaning up and normalisation of metadata should be suggested before the export takes place. Cf. certain tasks like standardising date format or certain values.

[15] Survey statement of Roxanne Wyns, KMKG

[16] Jef Malliet, Provincie Limburg.

[17] For example a thesaurus like the *Art and Architecture Thesaurus* managed by the Getty Institute.

IPR and licensing:

When managed locally, metadata is normally not subject to copyright protection and IPR and therefore is hardly ever considered when the metadata is made. When preparing metadata for publication, IPR and privacy issues become critical and need special attention. Digitisation of physical objects might entail new rights: on the metadata, on assets (the digital objects); on the thumbnails and on the database. The intended use and re-use of each of these parts have to be well considered from the start.

***Standardisation and normalisation*:**

Data structures are usually defined by standards describing how to register and organise metadata or information about objects within a single institution (e.g. the SPECTRUM standard). Contribution of this metadata to Europeana, or to an aggregator, tightens the requirements for standardisation because interoperability may not be considered as a criterion within these standards. Often there are several options in the recommendations, from which the collection manager is advised to select the best suitable for the collection. Such options will need to be more restricted in view of data exchange and aggregation, in particular for data elements that attract special attention and which are used to compare or relate objects to each other, e.g. the 'title' field, 'object type', and time period.

From physical object to digital asset:

On a local level, in the CMS, the object type usually refers to any type of object in the physical collection. The data structure of the CMS was designed and is used for describing physical objects.

A single physical object can be associated with several different digital assets like images, text documents, and audio files. However in Europeana each record is only allowed one representation. This is sometimes not clear to content providers, and so should be included in guidelines and specifications

Distinction and link between metadata and assets:

In most CMSs, the link between the database record and the attached digital assets (e.g. images and texts) is transparent to users. The digital assets themselves are not contained in the database and need to be managed separately according to certain rules of best practice. Attaching links to digital assets to database records and the proper archiving of the documents requires clear, accessible guidelines and the need to have a unique, persistent, identifier for each asset. Attention needs to be paid to:

- File size;
- File names (beware of special characters!);
- File types (and extensions);
- Quality (e.g. high or low resolution);
- Thumbnails - in case the asset is an image.

This lack of information shows that automatic transformation cannot always be done and interpretation will often be required. Some aggregators, as service providers, which are in closer connection with the content providers, can assist with this through appropriate support and direct customer advice. The ECK can help by providing guidelines and/or by pointing at best practices or examples. These will be useful for the aggregators as well, providing a uniform approach and stronger case. ECK users must be made aware of the implications and often need to make adjustments to their registration processes or registered data or even to the CMS functionality itself. Otherwise they will not be able to use the ECK in a satisfactory way, or may not be happy about the results.

3.1.2 Requirements for metadata quality and management

The transition from local metadata to global metadata requires collection holders to have a good grasp of the structure of their information and how it is modelled in their database (CMS). Therefore there are some criteria stated below which refer to best practise data management which should be possible with CMS systems. While these are strictly not requirements that concern the ECK, they involve possible additional steps that might have to be taken at the level of the CMS or whatever system is used to store the metadata.

It has been remarked, that even though partners often use the same CMS system, the way the system is used internally and how metadata is entered by the institution (e.g. description, field entry, date format, and thesauri) is unique, in some aspects, because vendors adapt systems to the client's needs. This makes the export of metadata from a database unique for each partner, often with lots of mapping conditions built in.

A very user friendly mapping tool where content providers can define which database field goes into which element according to a specific condition is definitely needed to allow flexibility. Non-IT specialists must be able to define what field value they want to map and export, into what element and what condition the value has to go - so change and enlarge the data export mapping without having to have an in-house developer.

Independently from which source the metadata is being transformed, there are certain requirements for CMS functionality to prepare metadata for further processing with the ECK:

- All data in the CMS must be exportable. Users must be fully in control of their database content and be able at any time to retrieve it independently of proprietary systems for exchange, aggregation, change of system, backup and long-term sustainability;
- It must be a possible to select which fields are to be exported or published on a record, collection or selection of records basis;
- Export to XML format is a requirement for good connectivity. It can: clearly indicate character encoding (UTF-8 is preferred); render complex data structures; be easily readable by machines and humans; be easily transformed into other formats by standardized transformation languages like XSLT or XQuery. It is the data exchange format in the Internet, and not more difficult to produce than any other format;
- The minimal export requirement is a format that faithfully expresses the data model in the CMS, as seen and experienced by the user, properly grouped and structured, with special attention to repeatable fields and segmented units of information;
- Export to standard data models should be available, preferably models that can render the richness of the data (e.g. LIDO)[18];

[18] ESE or EDM are not required as export formats from the CMS, but are requirements for the ECK.

- The ability to include 'constant values' when exporting metadata, e.g. all selected records have IPR value 'X', and data provider name is 'Y', and general possibility to change records 'in batch', e.g. enter the value 'X' in field 'Y' for all selected records, or include paths to thumbnails and assets.

3.2 Step 2: Select

The selection process entails the selection and management of sets of metadata records to be transferred to a target (e.g. Europeana or an aggregator).

Similar to the first step (Manage), the question is whether the select function should be part of the CMS or of the ECK. The majority of respondents see the advantage of this functionality being part of the CMS in having more control over the metadata because then the ECK does not have to access the whole database.

Detailed selection options are given below. Some of these will already be part of some CMS systems. Vendors can give indications of the extent to which they believe their solution meets this requirement already. Any gap can then be filled by ECK developments, although this might have to be solved on a 'vendor basis', especially when it comes to coding new technical solutions.

Users are often uncertain about what will happen to the information they write in their CMS. When registering information, users must be able to establish clearly and control which information will be passed on to which target and under what conditions.

Most CMS systems (as far as we discussed with the other technical partners at the 1st Network Event in Berlin) provide a method for locating a set of metadata eligible for including or excluding from Europeana with little or no development needs. These might be refined to provide more fine-grained selection of data if necessary.

A potential sub-step might be the scheduling of data exchange. Some institutions, for varied reasons, might want to restrict some of their metadata from being uploaded to Europeana before a certain date. Europeana inclusion set designations might need to include a date or timestamp for this.

Collection managers must be able to select what metadata they contribute to Europeana (or other target) in an easy and clear way, at various levels[19]:

- Manual selection: Hand-picked by record or by adding tags to selected records;
- Records selected on the basis of specific values in a variety of fields: e.g. by location, by object category, by theme, by section, or by (part of) inventory number;
- Composite filtering with clear Boolean operators;
- Selections for aggregators are not necessarily exclusive, and can overlap. If certain records are provided to more than one aggregator, they could arrive at Europeana by more than one channel (and in different formats);
- A collection holder must be able to manage multiple selections, for Europeana and for various aggregators. Selections can be based on different criteria and can overlap;
- Selected fields from all records, plus being able to filter according to specific content in some fields;

[19] Almost each of these selection mechanisms can vary according to the aggregator.

- Selection must be repeatable, e.g. at updates; filters must be storable and re-usable; it would be good if these filters can be understood also outside of the CMS, e.g. through the use of a uniform XML schema;
- It must be possible to exclude or include each individual digital asset attached to a record in a selection. It might be sufficient to indicate per digital asset if it can be published and reused or not;
- If there are many images attached to one record, it must be possible to select the one which will provide the thumbnail which represents the record, or to deliver more than one digital asset with the same repeated metadata;
- Users must be able to know, in a clear and user-friendly way, which metadata is included in a particular selection. This requirement must also be taken into account when the filtering of fields occurs as part of the mapping process;
- Whether certain records or fields are or will be included in a selection must be clearly visible and easily controllable for the CMS/ECK user.

3.3 Step 3: Prepare

Prepare the selected metadata for transfer to Europeana. Several sub steps occur in this phase:

- (Further) normalisation;
- Field mapping;
- Content mapping;
- Application of identifiers.

3.3.1 Metadata cross mapping

There are roughly two scenarios at this stage, depending on where preparation will take place:

Scenario 1: Connect ECK to CMS and prepare metadata locally

Some CMSs will provide the possibility to connect with or implement the ECK directly in the system. When this is the case, the metadata will be selected and prepared locally, in the CMS, before being transferred further to Europeana. An advantage of this scenario is that the steps taken and the objects shared will be documented in the CMS.

Scenario 2: Export metadata from the CMS and prepare externally

The selected metadata will be exported out of the CMS first before being further prepared for Europeana. This is for instance the case with platforms and aggregators where a tool like the Delving Platform[20] or the MINT tool[21] is used to map and then transform the imported metadata.[22] Here, the workflow might include the mapping to an intermediate format from which further mapping to the Europeana data format (now EDM) takes place. A reason for mapping to an intermediate format could for instance be that the national aggregator provides a thematic portal for the metadata as well, like the national Dutch aggregator for museum collections[23], which uses LIDO as intermediate format. The steps taken and the workflow progress will have to be documented in the ECK, since a connection with the source metadata in the CMS has not been established.

Even for institutions that can connect their CMS to the ECK, scenario 2 will be useful, e.g. for data coming out of other databases used in the same institution.

For the support of the objective of returning enriched content from aggregators and Europeana to content providers separate metadata cross mappings must be made for both directions.

Good cross-mapping requires very good understanding of the source and target data models by the user; e.g. when mapping metadata it is important to know what is meant by the source (intention of the data model plus interpretation by the registrar) and the target fields, so it can decided what the right content is for a field. The customisation of the data model is possible, but the implications must be clearly indicated to content providers. Equivalence and divergence of customised fields within standardised data models will seriously hinder the interoperability and should be avoided even when they are established and well documented. Content providers must become comfortable with their own data models, with the data exchange or Europeana data models applied, and with the relationship between both.

Mapping has to be done on a semantic level as well, not only on a generic level (see 3.3.2 ‘Enhancing metadata quality’). It would be useful to be able to save mapping templates for different usage scenarios or collection types. If a metadata cross mapping tool is created, it should be very simple to use. Users without much technical knowledge should be able to do a simple mapping of their data fields without having to consult their CMS vendor.

[20] Delving platform is or was used by the Dutch National museum aggregator. For the tool see: http://www.delving.eu/the-delving-platform

[21] MINT Mapping Tool http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mappings

[22] Both, the Delving platform and MINT mapping tool, should be reviewed and assessed for potential re-use or adaptation to be used with or as part of the ECK.

[23] National Dutch Museum Aggregator: Digitale Museale Collectie Nederland (DimCon). http://www.digitalecollectienederland.nl/

Other requirements suggested were:

- Users must be in control, and be able to see the result, e.g. what is transferred to where and in what form;
- Users (e.g. collection managers) must understand the source and the target data models, including any constraints (e.g. required elements). When specifications of target schema's change, the collection managers should be properly informed, and the changes should require minimum intervention from them and no extra costs;
- Standard mapping tables provided must be well documented and modifiable to meet specific source data model interpretations. These transformation style sheets must be coded in a standardized language (e.g. XSLT) and stored for re-use or exchange;
- Machine-generated transformation style sheets must be manually modifiable;
- For advanced users the mapping should take into account a wide variety of data structure issues:
    - Conditional mapping. The decision about which target field for some content may depend on the value in an attribute or in another element (or in a combination of attributes and/or elements);
    - Mapping must be able to take into account the structure of nested or grouped elements;
    - Values of some target elements may be concatenations of values in various elements or attributes at the source, and include constant parts;
    - Attention must be paid to character encoding (preferred is UTF-8) and special characters.

3.3.2 Enhancing metadata quality

Enhancing metadata quality requires tools and good guidelines for content providers. The role the ECK can play in enhancing metadata quality could be to provide these guidelines or to validate the delivered metadata and report back to the content provider if the metadata is not valid. In the latter case the content provider will have to update the non-valid metadata in the CMS and re-start the process. However, good data preparation starts with good quality of source metadata (formal quality, consistency and relevance). At this stage further normalisation could take place in order to prepare the metadata to be ingested by Europeana, such as:

- Inserting constant values in certain target fields (e.g. collection owner);
- Inserting derived values in certain target fields when source field is empty or contains metadata that is not valid (e.g. if the field ‘title’ is empty, then insert a value constructed from values in other fields;
- Remove certain unwanted values in certain fields (e.g. ‘unknown’ in the field ‘creator’ is semantically not correct and has no use).

Enhancing metadata quality and enriching the metadata at this point makes sense, because Europeana will only do an automated enrichment with uncertain quality. At this point the metadata Europeana receives is of variable quality, which makes enrichment more difficult. Enriching metadata is also something an aggregator could do, if the content provider prefers. For instance, to map on a semantic level it must be possible to map locally managed keywords to equivalent terms in broader standard thesauri (e.g. an institution’s own keyword list mapped to *Art and Architecture Thesaurus* terms or the own rights statements mapped to Europeana rights terms).

Possible functions include:

- It must be possible to remove particular words from specific fields: e.g. qualifiers such as 'ca.' or 'appr.', '?' to enhance the findability of the metadata, whereas these qualifiers are useful for presentation purposes;
- Provide clear guidelines regarding normalisation and standardisation of values at the registration stage. Any normalisation that needs to be implemented during the aggregation phases is likely to produce difficulties and irregularities;
- Define mandatory values (as prescribed by the schema used, or based on minimum quality guidelines);
- Define conditions (cases) for predefined string lengths, e.g. when [type] = "production place" THEN [eventType] = "Production");
- Define constant values, e.g. for metadata not included in the collection database, e.g. language of record, content provider name, Europeana rights terms;
- Define preferred and supplementary digital representations. Also, ECK can provide a tool that will generate thumbnail icons in the case the item "exported" to Europeana in not already online and therefore lacks one (and of course a PID for this thumbnail). What will have to be defined later, is where the thumbnail will be stored - in the ECK, the content provider's CMS, or in Europeana;
- If it did not happen earlier in the process, the metadata or media types must be defined here (e.g. thumbnail image and full size image) on record level or in batch.

3.3.3 Apply Persistent Identifier

In a collection management database a single physical object can be represented by several different digital assets like images, text documents, and audio files. These assets are usually not stored in the CMS itself. Assets which can be retrieved via Internet are stored on separate servers, apart from the CMS. There should be a URL or some other form of persistent identification constructed which points to the permanent location of the asset.

The ECK might have to consider the provision of PIDs also for items that are not available over the web or are only internally available through the ECK to support the process. For persistency it is preferred that the PID (URI) will be provided by the CMS and stored in Europeana or an aggregator, as it is the key for updates and deletions in Europeana from the CMS of the content provider. In case the PID is source of the PID is not the CMS the ECK will have to provide the list of published PIDs back to the institutions.

Other points of references like Institutions, Collections, Object-images and Linked concepts. The construction of the permanent identifier could be based on values in specific fields in the database, but the identifiers must be independent of the aggregator and persistent identifiers must remain constant for the content provider (may not vary according to the aggregator) to avoid duplicate entries. A solution for PIDs has been suggested by the HOPE project[24]. The re-usability of this work has to be investigated. Also, Europeana has set up a task force to define the needs of content providers regarding PIDs and the role aggregators like Europeana play to service them[25].

[24] https://pid.socialhistoryservices.org/

[25] http://pro.europeana.eu/web/network/europeana-tech/-/wiki/Main/Task+Force+Persistent+Identifiers

3.3.4 Apply License

Europeana rights guidelines[26] must be implemented. It should be considered that the rights for the physical objects might not be identical with the rights for the metadata and the assets (digital objects). Different sorts of IPR and/or licensing can apply to all three of them. It should be possible, if needed, to attach IPR information to:

- The digital objects;
- The metadata;
- The previews.

The question is whether the rights information will be part of the CMS action and therefore stored in the CMS or if this information will be added later by using the ECK, both are possible. There is a third scenario where the ECK must map the organization's proprietary licensing descriptions to the appropriate standardised copyright licensing schemes that adhere to the digital content and the metadata (e.g. Creative Commons.). Therefore, the ECK should provide functionalities to assist and guide its users to select the most relevant of existing licensing models and apply them where necessary, either in batch or to individual records.

3.4 Step 4: Validate transformation and receive feedback

Whatever the preparation steps and scenarios, wherever the metadata is prepared, in the CMS or in an external tool, for both scenarios it is important to have feedback about errors during the execution of the transformation operations. They must be properly reported so the user can easily find the cause, as they might not be an expert on data transformation. The result of this phase of the process must be shown to the user in an understandable way.

It is necessary to keep record of changes to the mappings in order to undo last changes (e.g. store last five mapping changes automatically or at least pose the question "do you want to save previous mapping"?). The ECK should provide a validation summary report to the end users at any point during the process and as well as best practises guidelines and encouragement to actively share knowledge with other users.

Possible requirements are:

- Clear feedback to the user / collection manager:
  - Which records did not pass the validation test?
  - For what reason?
  - In which field?
  - What was the content in the source data?
  - Which field is missing?
- Nice to have: 'what if' function, possibility to tentatively manually correct the input file in the place where it failed, without having to go through the whole export procedure, for troubleshooting purposes (once the error has been located and a solution found, the correction needs to be done in the source CMS);
- Ensure that the updated schema of EDM is being used for validation, including non schema-related requirements;

[26] Guidelines for the Europeana rights metadata element. See: http://pro.europeana.eu/c/document_library/get_file?uuid=06e63d96-0358-4be8-9422-d63df3218510&groupId=10602

- Make a clear list of all the requirements that need to be checked, possibly connected to the validation application, understandable for collection managers;
- Ensure that successful validation warrants validation also by Europeana at ingestion;
- Even without intended reuse, preview should be possible. It would be good to have a system that points to mistakes and gives tips on how to improve the mapping.

The content checker needs to be flexible to support data model changes. In the past the Europeana Content Checker was used as a network tool. Maybe for some of the validation requirements the ECK can build upon the knowledge and experiences of the previous Content Checker. Licence information must be verified here before it is sent to Europeana.

3.5 Step 5: Supply (push / pull)

Ingestion of metadata into Europeana at present can either be via[27]:

- Data pull by harvesting via OAI-PMH; the preferred method for Europeana at the moment. This means that Europeana has to take active steps to ingest the metadata;
- FTP/HTTP upload which should be considered possible at this point. However, it will require more manual work and has fewer features than OAI-PMH.

Automation of the data export workflow, at least in part, could improve the process. In most data export workflows, much manual work is needed. However, the experiences show that ingestion can be automated once Europeana has manually set the process to work.

The ECK will have to support both ingestion methods listed above with target and source functionalities. The implementation of an OAI-PMH server on the content provider side involves extra resources. Additionally, OAI-PMH does not provide interactivity or exchange of information about the status of the ingestion at all. If the use of OAI-PMH is intended, then clear instructions have to be given to content providers about what is needed. All the OAI-PMH options should be explained so that they can make an informed decision about their advantages and disadvantages, costs, and requirements for installation and maintenance.

An additional required method to the current two for data ingestion is a data push where the metadata is sent into Europeana without human intervention (like SWORD publishing which is more commonly used in the UK).

When it comes to ingestion, the delay within the whole process must be as short as possible. Because content providers want fast results, pushing might be the preferred method, but since the timing of import in Europeana is also determined by Europeana processes (e.g. indexing and enrichment) the time aspect is not really influenced by push or pull. The push method does allow the CMS suppliers to implement a solution that gives full control to the content provider. The content provider then determines what and when to upload to Europeana and which of the already provided records have to be deleted from Europeana.

There should be separate functions in the ECK for uploading, deleting and downloading enriched metadata. PIDs are used for object identification and need to be sent to Europeana if they already exist. Similarly the URIs or PIDs made by Europeana should be sent back to the content provider. This allows CMS suppliers to implement solutions that have direct results.

[27] http://pro.europeana.eu/web/guest/technical-requirements

Test ingestion should take place to give feedback about the validity of the metadata and to show whether the data provider has to take further steps to improve the metadata. Ideally the validation during this test ingestion should match the validation by Europeana. Metadata supply could also be possible via the built-in CMS functionality and have the following possibilities, which are known as 'incremental harvesting' and are a standard feature of the OAI-PMH protocol:

- Only transfer new records;
- Transfer new and updated records;
- Transfer all records.

Supply the metadata via an aggregator

As above, this is also possible in different ways via push or pull.

An additional 'configure' sub-step could be necessary if content providers want to connect to different aggregators.

The ECK should also facilitate collaboration with third parties and should provide the necessary tools (e.g. licensing filters and query APIs). A common services API set should be enforced by the ECK, following an Open, non proprietary approach.

3.6 Step 6: Data acceptance

After test ingestion, feedback should be provided to clarify whether the received metadata is valid for processing and storage in Europeana (the new environment).

One obvious prerequisite is the agreement by the content provider to the Europeana Data Exchange Agreement (DEA). If this agreement cannot be automated somehow by the ECK, it will be critical that the technical workflow of the ECK is aligned with the organisational workflow. Also at present the DEA is implemented at organisational level, while the exchange workflow should allow implementing license information at collection or even record level. Detailed log files or protocols are necessary in a format which the non-technical content provider can read and understand. As mentioned above, the validation during this test ingestion should ideally match the validation by Europeana.

The ECK could act as the communication channel about acceptance and any further steps performed at Europeana, like a kind of tracking system informing the content providers about the status of their metadata. The content provider must be informed timely about the conclusion of the final step of publication on the target website (Europeana or aggregator).

Content providers and aggregators regard the need for a proper visualisation of how the result will look in Europeana as helpful. At the moment Europeana has a web based content checker tool for that, which works by http upload of a small set of validated records. Technical project partners generally worry that the preview function could be difficult to realise, as Europeana may change its data model.

When metadata is accepted as being valid by Europeana, or other target, a confirmation message should be sent to the content provider and vice versa in case the data transfer failed.

3.7 Step 7: Enrich and Return (including Return and Re-use)

After the metadata has been provided to Europeana, or another target, enrichment of it might take place. Enriched metadata should be made available to collection owners through the ECK.

A service could be offered to content providers enabling them to browse their enriched metadata without having to implement them in their own CMS. This service could be web browser based and be used independently of their CMS. Collection owners should be able to use the enriched metadata in whatever way fits their needs best. They may want to include it with the metadata in their CMS or keep it apart, or even just include a pointer in a record that leads to enriched metadata.

The ECK should support the delivery of enriched metadata in various data formats. Since it has been suggested that the return steps are the same steps that also lead to uploading metadata, this requires metadata mapping in the other direction (from Europeana to content provider).

Europeana is already planning three ways to expose metadata plus enrichments: API, an OAI-PMH server, and Linked Open Data. Ideally the ECK re-ingest function should be built on top of these.

The return step could have the same sub-steps metadata has to take to be ingested into Europeana. Since there is still little to no experience with this workflow it will have to

- Manage (what metadata is available and in what form?);
- Select (which records or fields does the data provider want back?);
- Prepare (convert to other data formats?);
- Validate (does it match the data provider's requirements?),
- Re-ingest (update the CMS).

Other steps may be required, and other requirements applied to the steps. For example the data provider's CMS may need to be change to accommodate the returning enriched metadata, with a reference to its source enabling the quality of the information to be evaluated. Re-ingestion of metadata must be possible through a deduplication tool which exchanges information between two systems and checks for corresponding records.

A suggested minimum scenario includes a zip archive file handed down from the ECK to be handled further by the content provider and imported later in the CMS. But it should be mentioned that for most CMS's import data into the system is not a standard functionality available to users.

- Re-ingested enriched metadata must be at least linkable in the CMS, and clearly distinguished from their own metadata;
- When merging into the CMS there must be an option to accept or decline the metadata or selection of metadata, even by individual fields. User generated content might come in the form of remarks, suggestions, corrections, opinions, additions, selections, user-curated virtual collections. Each enrichment must be individually removable or flagged as inappropriate/irrelevant/fake/not-accepted by the content provider. Flags should be maintained locally and automatically applicable at a following update of the re-ingestion;

- The source of each type of enrichment must be clearly indicated, for example machine-generated (by which process?), added by an end-user, verified or not verified;
- Pull option, at the request of the data provider:
    - Immediate, delayed or according to a preset schedule;
    - Full or filtered: e.g. related to a specific object or group of objects.
- Push option, by means of a RSS-like solution:
        - At preset intervals preferably chosen by the data provider;
        - Triggered by some event;
        - Each time a new relevant item is added.

Additional IPR and, in the case of user-generated content, privacy issues will have to be considered (does the return of metadata to data providers include accepting a Europeana licence for the enhanced metadata?). Even if the metadata is public domain the user should accept that, so it is clear to the user what they can do with the metadata. In the future the licensing conditions may change.

The return workflow must not necessarily follow the same path as the uploading to an aggregator or to Europeana. Therefore, it is best described independently in a separate workflow.

Additional 'Store' Step:

This step adds extra possibilities like the comparison of the metadata that is sent back from Europeana with the metadata that was sent to Europeana. Other advantages of storing are the versioning of the metadata, so it is possible to keep track of updates. If the records are stored in a central system, like an aggregator system, it has the positive effect that the metadata from other institutions can be used as a best practice to map their metadata to EDM. If an institution has already mapped MARC21 to EDM, then other institutions should be able to reuse this mapping.

4 Business requirements

In chapter 2 we described some general assumptions and expectations of the ECK. Especially 2.5 dealt with business requirements of the ECK: what is required to make the ECK sustainable and reliably embedded in the Europeana value network. These serve as input for a future delivery on the forward plan. The previous section on functional software requirements looked at the workflow of exchanging cultural content and the requirements related to all the steps. Although the main objective of Europeana Inside is to automate or semi-automate the submission of content from CMSs and aggregators into Europeana, developing software is just part of the solution. The lasting transformation, the mission of Europeana Inside cannot be achieved by technology alone. That is why the project proposal not only describes technical barriers but also organisational, legal and financial ones.

This section looks at the process of providing content through the ECK from an organisational, institutional or 'business' point of view. It is important to stress that business is not merely used in terms of economic value creation but refers to wider organisational processes, including more fundamental or strategic business conditions for a successful implementation, and more applied or operational conditions. The range of functionality described below form a framework in which an efficient implementation is possible. As such, they are related to the context of use, but unlike the context of use and its holistic value network perspective, this section looks at business requirements or conditions within partner organisations.

While many of these requirements are not directly within the scope of the ECK, they are important background conditions. Potential users of the ECK will have to face them and optimal conditions serve as a prerequisite for the actual use of the ECK. Also the sustainability of the ECK in the long term depends on the Europeana Network together somehow finding solutions to the business requirements that satisfy all interests. A knowledge base filled with results from research and best practices on digital heritage would be a useful companion to the ECK. Europeana Inside can investigate in WP4 what is actually most needed and then decide if an extra deliverable will be part of WP5.

4.1 Strategic planning

It is usually part of a content providing institution's mission to make its collections publicly available, and to provide information on them with multiple interpretations of the material. Allowing access to this cultural content via Europeana and other portals is a strategic decision to be taken by the institution's board. It must bring added value to the institution, which can be social, cultural or economical. Most institutions focus on the growth of added social and cultural value, encouraging people to be acquainted with aspects of culture heritage.

Adding economic value creates more difficulties. Investment is needed to be able to deliver access via Europeana, or any other portal, and should at least be balanced by more, for example: visitors to their website and online catalogue; physical visits to the collections; publicity; enriched content.

Many content providers deliver metadata to Europeana already but usually through participation in a project that brings extra financial resources. To them the ECK lowers the barrier to continue with the exchange of more and richer cultural content, possibly also via different channels.

It is also the intention of Europeana Inside to reach the vendor's broader user-base of institutions, including many that do not yet contribute to Europeana. These new participants first need to be convinced of the added value of delivering metadata to Europeana. Only when the added value is clear, will more institutions make the strategic choice of participating.

For technical partners, an added value would be to certify a CMS as 'Europeana ready' or 'ECK compliant', giving them an advantage over competitors who do not have it. Accreditation for CMS suppliers will help promote Europeana and stimulate also CMS suppliers that are not partner in Europeana Inside, to implement the ECK.

Unique selling points for the ECK are:

- Exploitation of existing web presence management software infrastructures of the participating content providing Institutions. These include existing web content management systems and their interfaces with the internally operated CMSs;
- Provision of supporting browser-based tools that can operate in parallel with the existing CMSs. In this way, groups of functionalities such as enriched metadata feedback from the aggregator could be smoothly assimilated in the existing workflow of the curators' tasks and activities;
- Consideration of copyright licensing information in addition to the Linked Open Data scheme for metadata.

Together with the overall benefit of the ECK of efficiency and effectiveness to the exchange process, these unique selling points will make sharing of cultural content more attractive as digital strategy to organisations within the European cultural eco-system.

4.2 Quality control

The quality of the information exchanged is crucial. Some kind of quality framework must be in place when exchange of cultural content takes place to promote high quality and monitor the performances of stakeholders in the value network. If the metadata can be normalised, perhaps even annotated in batch, and improved (e.g. adding references to external sources), this would be beneficial for:

- Europeana;
- The end user (viewing richer content);
- The content provider (more and for a longer time span).

If the metadata are of poor quality then retrieval is difficult or even impossible. Therefore every enhancement has to be made very carefully and needs to be identifiable as such. Good quality should not only be supported but guaranteed.

A discussion on the quality standards may be held within the consortium to establish guidelines about the quality control, but this must be coordinated with the quality discussions in the overall context of Europeana. There are several places where quality control can be applied. These include in the collection management system before and after exchanging metadata with the target, Europeana or another portal including the validation steps. This must be foreseen in the operational planning and infrastructure.

Establishing quality standards must be one of the requirements taken into account during the process of exchanging. For a content provider the definition of high quality information can depend on the quality standards attached by the organization's priorities or by the interpretational context. Also some metadata will not be included in a CMS because it is project specific or it is just not completed in the CMS record. Often no batch changes are possible, only to manually add information to each record. Therefore, there should be guidelines formulated on making good, exchangeable, metadata, and guidelines for creating metadata that can be more easily enhanced.

From a technical perspective, it would be beneficial if the ECK served as a step for standardised quality control of the metadata to be published. It should be a tool for promoting quality enhancement of metadata. The ECK validation of the metadata should lead to some form of certification or protocol for cooperation with supporting aggregators and Europeana. For example, such a protocol might give priority to ECK validated metadata or could use the copyright licensing information of the annotated digital artefacts to allow for third-party controlled and licensed usage. The feedback from the quality control has to be provided to content providers in a sensible way, in order to be used for defining a strategy for data quality improvement.

4.3 Business model development

All stakeholders participate in the exchange of cultural content because it brings them added value. Business models are a way to discuss and optimise this added value. Business modelling approaches can look at the whole stakeholders configuration of the ECK (see Section 2: *the value network*), or they are analysed from an individual organisation's point of view. This section will focus on the latter.

Added-value services and content that may be realised through cooperation with commercial third-parties could strengthen the business model of content providers working with the ECK. In this mode, copyright licensing information must be available as part of the metadata. This allows for promotion and possible exploitation of the actual digital content in a controlled and secure way. Furthermore, the enabling of business-driven use cases directly through the ECK, or at least as an additional distinct mode of operation, could be considered.

Europeana needs the contributions from data providers and aggregators. The focus of the ECK is on getting metadata more efficiently in Europeana, and on getting enriched metadata back from Europeana. The result is a stronger business model for Europeana in relation to content providers. In the strictest sense, it could limit itself to Europeana data formats as output. However, then it may not be sufficiently useful for aggregators, who often have additional business goals that require different data models, as they may not merely pass metadata on to Europeana. In that case, it will not be optimally useful for data providers either, since Europeana prefers direct delivery of metadata from institutions to be an exception and prefers to work with aggregators. As such, added value for aggregators and content providers is directly intertwined.

Heritage institutions are not usually familiar with business model administration and most content providing institutions operate with public funding, which require different business models. In developing these models there should be a discussion about the application of particular business models on sectors of activities (e.g. education and shop). Using the knowledge and experiences already available to a common discussion an optimization of the added, social and economic, value would be useful. Depending on what the benefit is different business models can be developed to support this.

For internal use especially, these models should be simple, to the point and catchy. Europeana Inside has to be able to convince management boards with a single page of information on the value of both Europeana and the tools that are needed to participate without having to have too much background knowledge of the process.

Examples of added value by the ECK that can strengthen institutions' business models are:

- Links to 'other sources' based on the enriched metadata which comes back from Europeana. This would be a real incentive to take part in Europeana. Examples include "Show more from this Artist", "Time", or "Location". Which museums hold works from this Artist, Time, or Location? A list should be defined on metadata fields and functions, which can be used together with the Europeana API to provide to every content provider "as a reward" and added value for taking part in Europeana;
- The exchange of cultural content represents an opportunity for promoting the cultural content of heritage institutions. The added value is brought by the interactive relationships with different audiences that are created;
- Increased visibility on the web will help promote the content provider. The enrichment of metadata, especially if this is also in the area of vocabulary control, should enable the content provider to enhance the quality of their own metadata. In Europeana, the content provider will see their metadata in the context with the metadata of other content providers. This value increases with more content providers participating.

In the business models of most stakeholders, e.g. data providers, service providers, and aggregators, delivering metadata to Europeana will not be their core business. Europeana may be important as a by-product, which brings additional resources, additional benefits, and enhancements to the core business. However data providers and aggregators have requirements and data models that have to support these other purposes, for which the Europeana requirements and data models are not designed. This should be taken into account when developing an ECK business model.

4.4 Managing statistics

Keeping control of the process requires quantitative information. The gathering of quantitative information will be useful not only to control but also to anticipate adjustments and future needs on processes and tasks related with the exchanging process. On the other hand, statistics also help the heritage institution to gather information about its own role in the project and afterwards about its role in the overall ecosystem.

Content providing institutions have a need for various quantitative statistics, such as:

- The number of records describing items that have been exported since a particular date;
- The number of records not been exported;
- Which items from which collections;
- The quality of the exported items;
- Which records have been updated since a particular date;
- Information about the user impact of exchanged metadata compared with that of non-exchanged material;
- Quantitative information on the aggregation of exchanged records coming from the heritage institution as well as the percentage of their participation in the project;
- Quantitative analysis of social reactions or feedback;
- Information on reuse of metadata by third parties.

The content provider needs to have the tools to get quantitative information from their CMS. The various content management systems together with the ECK should provide these functions and manage the related metadata. Statistics might also be a requirement for some content providers in order to allow external auditing by a public funding organisation.

4.5 Operational planning

The working process of exchanging cultural metadata needs to be embedded in the organisation. This requires a clear definition of tasks, authorities and responsibilities across all stakeholders involved. In the institution operational planning also must be clearly defined across various departments that are involved with digital heritage and service, even though few content providers may have experience in doing this.

Operational planning also needs to take into account the need for flexibility. It is required to monitor the business process, report on it, and adjust it when necessary. The monitoring process needs to be a part of the definition of tasks and responsibilities.

However operational planning itself will be outside the scope of the ECK because this is an organisational issue. The ECK can support some of this planning by providing tools, guidelines and documentation for user management, defining different rights to different users, and automatic messaging for users.

Some, especially larger, content providers have participated in several Europeana-related projects, or have information accessible through several, sometimes overlapping, aggregators. They need to be able to keep track accurately of what metadata they have contributed, including updates, to Europeana through which project or aggregator, each with their own requirements and processes. The ECK can be a valuable tool in doing so. It is particularly important that the ECK procedures are well documented within an organisation to cope with staff changes and retraining. It is also suggested that the ECK should be embedded into an open source workflow engine.

4.6 Return on investment

Innovations, like implementing new software and making new partnerships, cost money. As stated earlier, in order to remove barriers to participation and to enable more institutions include the decision to provide metadata to Europeana in their strategic planning, it is necessary to clearly demonstrate what the benefits are. The return on investment is linked to business models, but is worth discussing investment separately to put more focus on it.

There has to be a clear idea about the balance between costs and benefits in the long term. Making new partnerships may be included in the long term social mission of cultural heritage institutions as it generally constitutes of a contribution to the world wide cultural interplay.

Usually cooperation, collaboration and partnerships are not prohibitively expensive. Implementing new software, however, costs a lot and it may seem very difficult to have a clear idea about the long term benefits. This is because technological platforms are rapidly changing, and exact measurement is hard. Financial data is needed to quantify the economic benefits, but it will be too farfetched, in the opinion of some partners, to predict and monitor the financial benefits of the ECK.

Small institutions have difficulties in investing in new software and new partnerships. So the future benefits and advantages must be clearly defined to support the decision of participating.

The ECK might stimulate Institutions to invest in using the existing software that supports it instead of developing new software. There should also be a free reference implementation that can be used by organisations that use a custom database and want to contribute their metadata to Europeana.

Publishing metadata on an institutional portal, using a direct flow from the CMS to the online catalogue, is not the same as sharing it with others like Europeana. It requires additional investment to export, harmonise and deliver the metadata according to the rules of the target. It requires investment in technology (e.g. data export system and OAI-PMH implementation) and human resources (e.g. mapping, harmonisation, and contact management). As a minimum, good connectivity requires more stringent application of the standards and rules for the information contained in the databases. It may also require some additional units of information (fields), and the availability of an XML data export function.

However, these are not such heavy investments. They should be also be part of standard functionality for a quality CMS in our information age, and not an optional add-on only available at additional cost. The return on investment can be envisaged as the numerous new opportunities for connectivity that an updates system enables.

The ECK can deal with further requirements, such as mapping functionality, and guidelines and functions that should be available across systems. A return on investment is certainly needed (e.g. more visitors to an organisation’s online catalogue, museum visitors, orders for photographs, enriched metadata returned and manageable in own environment), however stakeholders can expect few benefits to be financial.

It is important to note that the ECK should remove barriers for the contributors to Europeana and not create new barriers. A barrier that could appear in the ECK is the extra cost for buying, licensing new software, or the need to buy add-ons. This should be avoided by creating an open source reference implementation, which can be used as an alternative. It is important that this implementation does not compete with existing implementations. The advantage for the vendors is that they can enhance the ECK and make the integration with the ECK and Europeana better and easier to use than the reference implementation. They can also support organisations with mapping and exporting metadata to Europeana.

4.7 Risk management

Being engaged in innovative activities brings risks, especially if the activities include interdependencies between different organizations. Assessing the risk and planning for counter management prevents serious damage. Like statistics, a good risk management plan is a management tool that should be applied. It helps the success of any endeavour. To avoid and prevent risks contributes to a positive outcome.

Extra attention should go to risk management of licensing issues, like the easy and quick deletion of records or fields. Risks are also higher when closed proprietary systems are used with limited connectivity. When changes are required, or requirements change, a collection manager can be faced with serious loss on investment, in particular for the time required to reconstruct the information contained in a database. Therefore a basic requirement for connectivity, an XML export, will help reduce risks. Although technical partners stress that risk management falls outside the scope of the ECK proper documentation on how to handle risks involved with exchange of collection data could be considered a requirement.

4.8 Contract management

Contracts between partners should be used to define the relationship and so minimize risks. In a project which integrates several institutions it is crucial that the work between partners be defined in a contract. Institutions should be aware that their own work process can affect those of the other participants. However the project is, at least currently, not concrete enough in terms of what needs to be done and who is responsible for what, to put this in a contract. This can only be done when there is a very clear agreement on what the responsibilities are. A firm engagement for maintaining vital parts of the ECK would be required, e.g. Europeana would need to be responsible for keeping the ECK updated whenever requirements change, data models are adjusted or new functions are created.

It should be noted, that Europeana already put together the Europeana Data Exchange Agreement. Every content provider wanting to participate in the Europeana aggregation infrastructure is required to sign this document.

The process of administrating rights and obligations between content providers and re-users is further complicated by the fact that aggregation often encompasses several steps through sub-aggregators subject to national regulations. The complexity is expected to grow because content providers want to differentiate the rights of reuse according to the collection (or even records and fields) involved and the purpose of reuse. The use of common licensing frameworks like Creative Commons will diminish complexity though.

Further input on this topic is required. However, some partners feel that contract management is important but beyond the project objectives.

4.9 Data modelling

Knowledge of the details of collections management practise in an organisation is important when mapping metadata to a general metadata standard to make crosswalks into another context of use, like Europeana. Data modelling must take into account how cultural content is collected and managed in different institutions and different collection management systems and how these can ingest and disseminate metadata in a different context. In practice most content providers have established various ways of dealing with their metadata. Interoperability is important, but difficult to achieve because curators and people dealing with data exchange think differently and are not necessarily aware of the other's needs even within the same institution.

A solution would be to use the services of an aggregator. It is the responsibility of the aggregator for doing the data modelling and describing the model in a way that the data providers understand. Of course if the data provider has the know-how for this, they can do it themselves. Data providers should be aware that any custom interpretation of applied data models has consequences on connectivity and exchangeability of their metadata.

An important achievement within the Europeana Network over the last few years is the development of the Europeana Data Model (EDM). While up to then aggregators worked with the minimum common denominator (usually based on Dublin Core), EDM is meant as a universal exchange model for the cultural domain and even beyond. It is also Linked Open Data compliant which makes it future proof.

However, at the moment, there are some issues with its use:

- It is complex and understood by just a few specialists in the heritage domain;
- It is not a stable data model yet;
- It is not designed as a universal data exchange standard, but for Europeana purposes only;
- Some important areas are not covered (mainly events that are covered by LIDO);
- Further crosswalks from traditional domain standards like EAD, MODS and LIDO still need a lot of effort from the communities (intellectual and compromises);
- Its implementation into Europeana is still at its early stages.

However, because of its use by Europeana, the Europeana Inside project should support its development and enable its implementation.

4.10 Partner management

The establishment and maintenance of a sound working relationship between partners is important. This includes the processes of finding the right partner and making agreements that are clear. New partners should share the same vision and priorities on how to provide access to cultural content. The public benefit should be central. Although partner management is not a core activity of the ECK and partly outside the objectives of the project, partner management can be supported by creating best practices or guidelines.

There are various ways in which Europeana in general and the ECK in particular can contribute to the establishment of new partnerships:

- ***Knowledge sharing and best practices.*** Working together has the advantage of knowledge sharing and learning from the mistakes and success from other organisations. There should be opportunities for knowledge sharing. This makes possible the sharing of interpretations of the Europeana Data Model, and striving for a uniform model in Europeana, or interpretation of the Europeana models.
- ***Long term commitment****.* The ECK should be supported by a community that will keep the ECK alive after the EC-funded Europeana Inside project ends. During the development of the ECK life after Europeana Inside should be taken into account. The vendors, aggregators and Europeana must agree to work together to continue the development of the ECK. There should be a development plan giving details about which features will be added after the project, who will contribute, and what the financial structure will be. Usage of the ECK should be monitored and evaluated. Also vendors and developers not in the project should be able to contribute.

This last business requirement enables this deliverable to come full circle. Partner management is, to some extent, an individual stakeholder's perspective on the important value network issues described in the context of use at the beginning of the deliverable. While the previous required business conditions depart mainly from individual stakeholders' situations, partner management as an important business requirement makes it clear once more that the value network balances need to be kept in mind at all times.

All relevant stakeholders, not only in the project but in the whole Europeana value network, can fulfil their own business requirements, e.g. their own strategic goals and planning, their own business models, and their own return on investment. Partner management from a stakeholder's point of view facilitates the stakeholder to reach its business goals and requirements. Partner management from the whole value network perspective refers to the balancing act that reconciling all these business goals might entail. This must continuously be kept in mind when working out both functional software requirements and business requirements that are desirable for all stakeholders.

5 Conclusions and recommendations

5.1 Results

This deliverable offers a broad overview of the requirement for ECK:

- What it is;
- What its application area is;
- What benefits it should bring;
- How it should function;
- What it is supposed to support regarding the workflow of exchanging cultural content data;
- What its relation is to other working areas of heritage institutions ranging from strategic to operational.

The description is based on the knowledge and experience of various representatives from the four stakeholder types within the fast developing European digital cultural ecosystem. Because the development's impact on the role of different stakeholders within the system remains unclear, requirements may differ depending on the stakeholder's perspective.

Because of the complexity of the subject, and the fact that a lot of the development is still in early stages, it is inevitable that the description offered here is incomplete, contradictory, and in the future will probably prove to be inaccurate in some points. However it does represent the expectations of the various stakeholders in this early stage of the project. Within this dynamic context, it offers an important milestone.

Here are some general conclusions about the ECK:

- The context of use of the ECK, i.e. contributing to Europeana aggregation is only a short term goal. The long term goal is a continuous participation in the digital cultural ecosystem in Europe;
- Specifying ECK requirements, we will be confronted with different existing scenarios by which content providers exchange cultural content;
- The ECK seems to simplify the connection between sources (collection management) and targets (collection users). For the moment we consider that Europeana and aggregators play a similar role, as potential target. The same applies to content providers and aggregators as potential sources.

  It remains to be seen if this model is static. Possibly we will end up with a more complex situation regarding the position of aggregators and the roles they can play in the ecosystem. Is it mere administrative or is there also an active role regarding the quality of the metadata, the uploading process? One thing is clear. They are a corner stone in Europeana's current strategy;
- Most expectations regarding the functionality of the ECK are shared by all partners. Main differences apply to the context of use:
  - Vendors wish to focus on Europeana as context of use while content providers have an interest in a broader application;
  - Content providers expect no extra costs for the exchange of cultural content while vendors fear extra unpaid technical support.

- For gathering functional requirements for the ECK a tentative workflow is described. It consists of seven steps: manage, select, prepare, validate, supply, accept, enrich and return. This process was somewhat confused by implementation issues. The ECK connects with the collection management systems on the one side (source) and Europeana or aggregators on the other (target). Duplication across systems should be avoided while at the same time dependency of the exchange process on commercial products or project infrastructure is a risk for the success of Europeana Inside in the long run;
- The exact relation between the ECK and CMSs is not determined yet. Will the ECK be part of the existing collection management systems? Will it be part of a new collection management system? Or will it work as a web-based stand-alone application? Partners indicated they prefer to use their existing collection management systems in some way, but if that is the case, a number of questions are raised:
    - Which collection management systems will be compatible?
    - Are there adjustments needed to the collection management systems, or will the ECK be designed to work with all systems? An inventory of existing systems and their specifications may be needed as well as a list of specifications that are essential in order to make the ECK work with a CMS.
- For any processing between the source and target systems the question is: Should this be statefull or stateless, and should the exchanged content be stored somewhere in some kind of intermediate format? The answer to this question has major implication for the future exploitation of the ECK;
- While the route to metadata ingestion into Europeana has been covered in the functional software requirements, there is not yet a lot of attention for the second function of the ECK, which is enabling the re-use of the enriched Europeana metadata by the content providers as well as by third parties. It is probably too early to define requirements for that, since it is not entirely clear exactly how the metadata will be enriched. Also the degree to which content providers would like to re-use their own enriched metadata might vary. Not everyone would like automatic reverse updating of certain fields in their CMS;
- The decisions on specification and development of the ECK might have major implications for its sustainability. However it is very difficult at this stage to determine the costs and benefits of the ECK: central versus distributed, commercial versus public, projects versus infrastructure, European versus national. The exploitation of the ECK is not only a matter of administrating, servicing and maintaining (open source) software. It also has to do with keeping the value network together and the knowledge context up to date.

Some general conclusions about the stakeholders:

- Only part of the management, exchange and reuse of cultural content is technical and can be improved by technology (i.e. software). Effective use of the software depends heavily on knowledge and understanding especially at content providers. The section on business requirements acknowledges this. Technical project results have to be specified and developed with the business requirements in mind, although fulfilment of business requirements is out of scope of the ECK and some partners even say outside the scope of the Europeana Inside project;

- Taking into account the complexity of the existing value network, the ECK should be of benefit to all four types of Europeana Inside stakeholders: content providers, aggregators, CMS vendors and Europeana. The individual content providers are the foundation for this network. At the moment their ICT infrastructures forms a main barrier to take on the role of network organisation. Fundamental changes to the working procedures, staff expertise and IT architecture are required. Europeana Inside can stimulate this transition. Content providers and CMS vendors are bound together in facing these challenges;
- The traditional business model of software vendors might be a barrier to the ambition of Europeana Inside of realising a major transformation of the number and scope of institutions participating in Europeana without extra costs. The search is for a future where the commercial market of digital heritage has a cost-effective and value adding influence on the public infrastructure for digital heritage;
- Even when the technical exchange of cultural metadata turns out to be stateless, it is clear that the organisations that play various roles in the process will somehow be linked together. The impact of this collaboration, its governance and how to manage the value network is an important success criterion, but it is a relatively unexplored topic that might bring additional ECK requirements. At the moment content providers stick to the idea of remaining in control over what happens to their metadata;
- Exchange of digital cultural content is a new area of expertise combining several traditional expertises, some of them considered alien to heritage professionals. There is a difference of opinion regarding how technical the heritage professional of the future should be. Data modelling seems to be at the centre of the debate of new skills required. The less the profession of heritage evolves, the more institutions will depend on technical partners, which are sometimes public but mostly commercial;
- Successful exchange of cultural content starts at the source. Attention paid by content providers to metadata quality, data curation, generic work processes and technical architecture, and consistent data modelling lessens the gap that the ECK needs to bridge. Heritage institutions need to improve their performance on all aspects mentioned above in order to make the vision underlying Europeana Inside work.

5.2 Impact

Based on the information brought together in this document the project can discuss and prioritise requirements, keeping the balance between improving the ecosystem without disrupting too much of what is at stake for the different types of project partners. The subcontracted task on legal aspects of Europeana Inside will add further requirements.

The next step will be to select use cases that can help specify in more detail what functionality can fulfil the main requirements of exchanging cultural content. The use cases will focus on what can be identified as the generic scenario. Simultaneously alternatives have to be analysed to see how much of them can be supported by the Europeana Inside results.

The *Functional Requirements* (D2.4) will provide the project with a comprehensive set of functionality described in enough detail to act as a bridge between content providers and technical partners making clear what can be developed into software in order to fulfil the ambition of Europeana Inside to simplify as much as possible to participation of cultural institutions in exchanging content as in Europeana. Functional requirements that make it into D2.4 will also be accompanied by quality criteria. These can be used in WP4 to assess if the software delivered in WP3 meets the specific requirements in order to satisfy the users.

At the final stage of WP2 the technical specifications have to be produced. Also for this deliverable the current document is an important source. It will form the basis for the development work in WP3 and WP5 and for assessing what part of the functional specifications will be covered by software development by Collection Management System vendors or by Europeana. The remaining parts that will not be taken on board there will have to be dropped from the project or, if possible, be covered by other tools that can complement the ECK.

References

Ballon, P. (2009). *Control and Value in Mobile Communications. A political economy of the reconfiguration of business models in the European mobile industry. (PhD thesis.)* Brussel: IBBT-SMIT Vrije Universiteit Brussel.

Ballon, P. (2007). Business Modelling Revisited: The Configuration of Control and Value. *Info - The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media, 9*(5), pp. 16-19.

European Commission. (2011). *Commission Recommendation of 27.10.2011 on the digitisation and online accessibility of cultural material and digital preservation. C(2011) 7579 final.* Brussels: European Commission.

Europeana (2011). *Europeana Strategic Plan 2011-2015.* The Hague: Europeana Foundation.

Europeana Inside (2012). *Europeana Inside Description of Work, version date 2012-02-22.* (Internal project document.)

Faber, E., Ballon, P., Bouwman, H., Haaker, T., Rietkerk, O. & Steen, M. (2003). Designing business models for mobile ICTservices. In *16th Bled E-commerce Conference, 9-11 June 2003, Proceedings.* Bled, Slovenia.

Friberg, A., Smith, D., & Brunelle, L. (2012). *D2.1: Europeana Partner Strategy and Development Plan. July 2012. Deliverable of the project Europeana Version 2.* The Hague: Europeana Foundation.

Osterwalder, A., & Pigneur, Y. (2009). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers.* Hoboken, NJ: John Wiley & Sons.

Van Passel, E. (2011a). *Valorisatiemodellen voor onderwijs, onderzoek en culturele industrieën (Archipel WP2, Deeltaak 2.1).* Flanders: Consortium Archipel. Available (in Dutch) at: http://archipel-project.be/sites/default/files/D2.1_Valorisatiemodellen_v1.0.pdf (Accessed 17 August 2012).

Van Passel, E. (2011b). *Een strategische roadmap voor duurzame archivering en ontsluiting (Archipel WP2, Deeltaak 2.4).* Flanders: Consortium Archipel. Available (in Dutch) at: http://archipel-project.be/sites/default/files/D2.4_Strategische_roadmap_v1.0.pdf (Accessed 17 August 2012).

Verwayen, H., & Limonard, S. (2010). Business Model Innovation. In: *Business Model Innovation Cultural Heritage.* The Hague: DEN / Kennisland / OCW, pp. 6-21. http://www.den.nl/art/uploas/files/Publicaties/BusModIn_eng_final.pdf

Appendix 1: Suggestions about ECK

This appendix consists of two figures provided by the project partner Zetcom. They suggested that the ECK should consist of two parts:

- ECK local (the connector to the CMS);
- ECK global (a web service that offers validation and transformation.

If aggregators are actively using the ECK, they should be regarded as content providers. This would also be the case, when aggregators enrich the metadata they receive from the content providers.

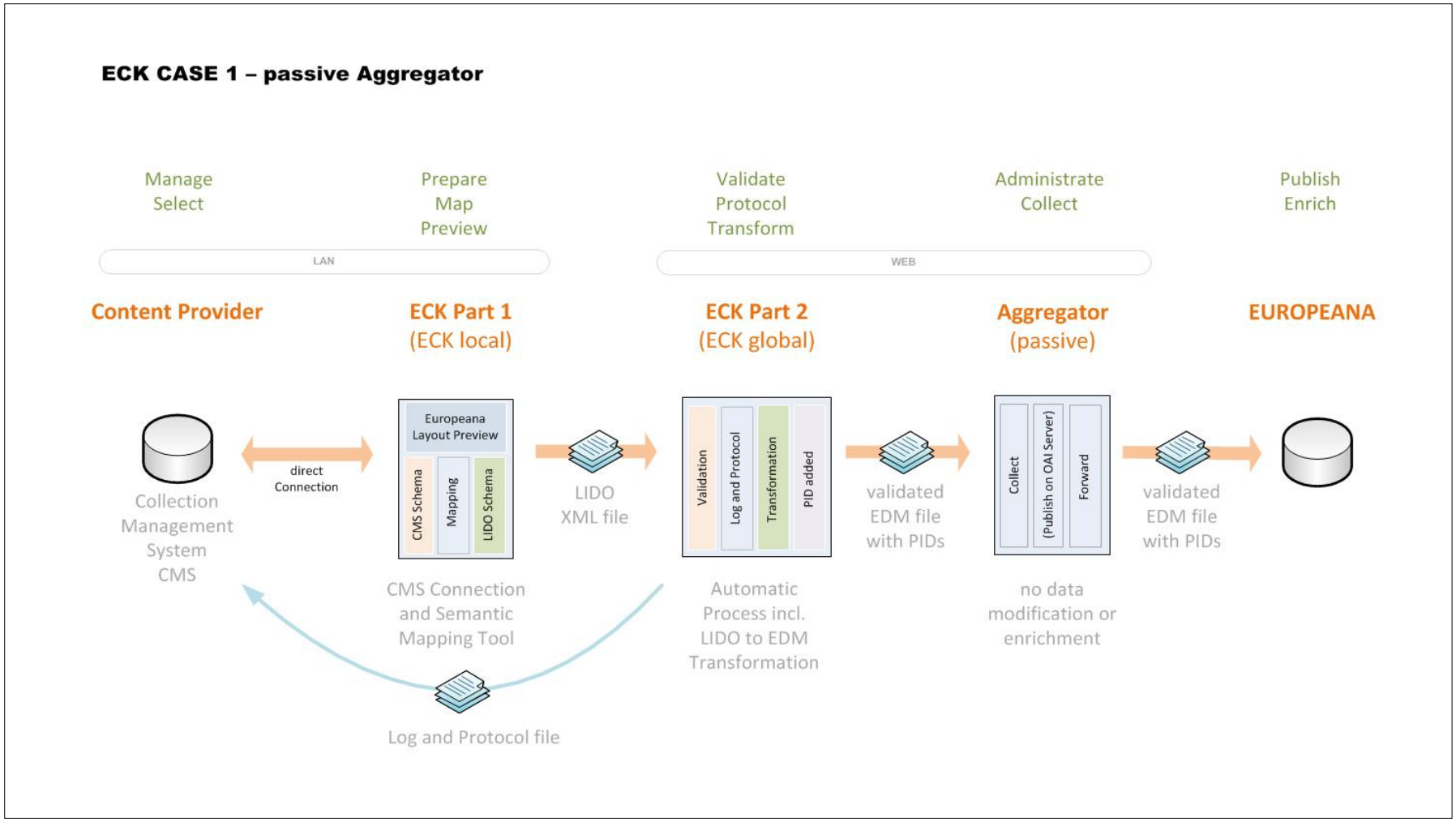

Figure 3 – ECK CASE 1: passive Aggregator

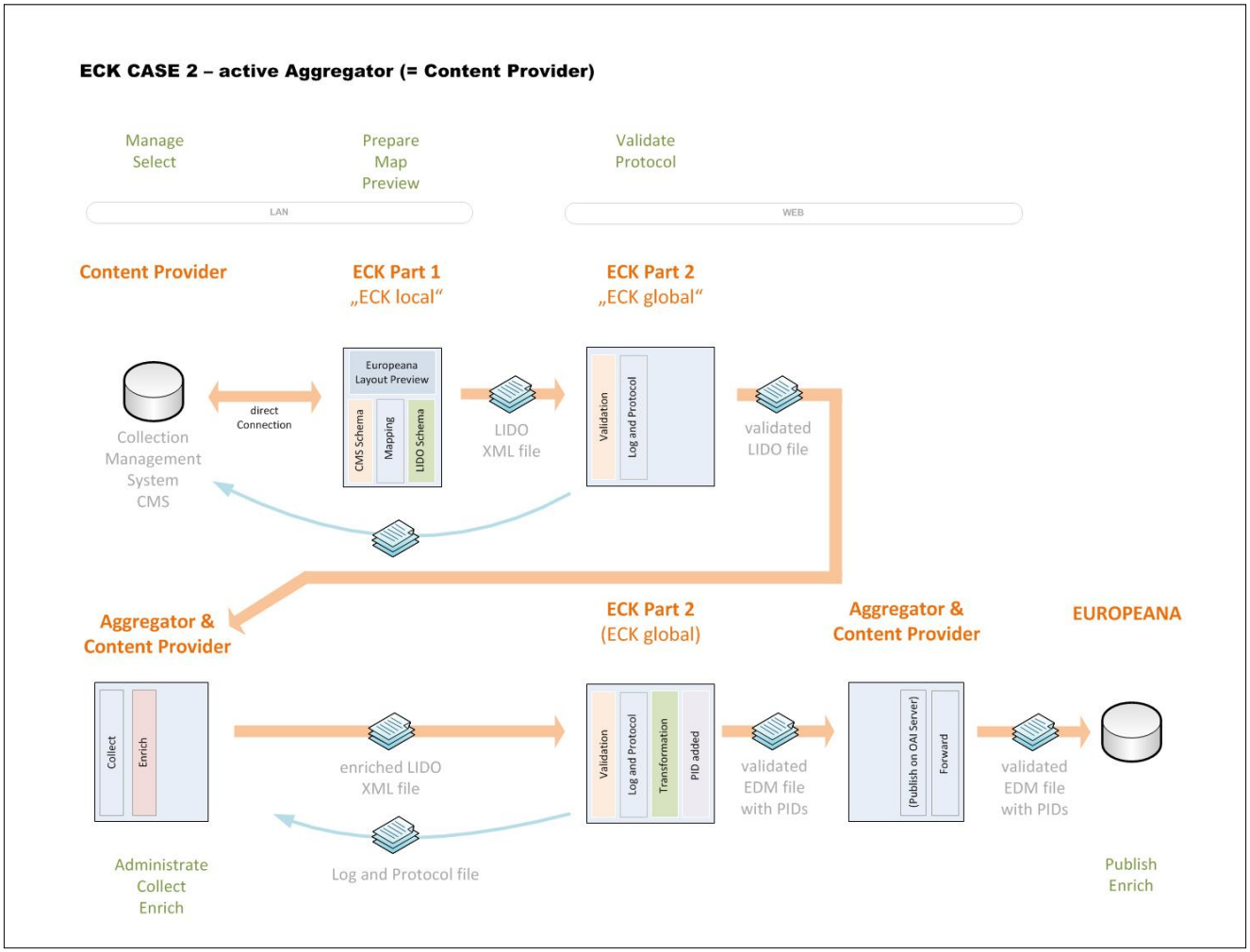

Figure 4 – ECK CASE 2: active Aggregator (= Content Provider)

Appendix 2: List of Task Assignment Contributors

Below is a list of people who have been put forward by their respective organisations as having participated significantly in completing the task assignment for each organisation, on which a large part of this deliverable is based.

Adlib Software Ltd (ADLIB)
Rene van den Heuvel

Belgian Institute of Natural Sciences (RBINS)
Laurent Meese

Benaki Museum (BEN)
Zoe Metaxiotou, Ifigenia Dionissiadou, Chrysoula Kondaki

Catholic University of Leuven (KUL)
Sam Alloing, Luc Schokkaert, Diewer van der Meijden, Luc Lannoy

Collections Trust (CT)
Gordon McKenna

Hungarian National Museum (HNM)
Szilárd Goda, Andrea Szebeni

Interdisciplinary Institute For Broadband Technology (IBBT)
Eva van Passel

KE Software Ltd. (KE)
Alex Fell, Vincent Bodinier, Ben Sullivan

Knowledge Integration Ltd (K-INT)
Neil Smith

Länsmuseet Västernorrland (SLV)
Bengt Wittgren

Mobydoc (MOB)
Eric de Cacqueray, Jean-Yves Cueille, Vincent Leconte, Jérôme Berdeaux

Monguz (MON)
Ágoston Berger, János Pancza

Município do Seixal (SEI)
Fernanda Ferreira

Museum of Fine Arts Budapest (FAB)
Krisztián Fonyódi, Ádám Pogány

National Gallery - Alexandros Soutzos Museum (NAG)
Elpiniki Meintani, M. Lambraki Plaka, O. Mezafou, L. Tsikouta, A. Giannoudaki, A. Zervou, E. Agathonikou, M. Kasimati

Petőfi Literary Museum (PIM)
Zsolt Bánki

PostScriptum Information Architecture Ltd. (PS)
George Koutalieris, Vassiliki Psatha, Spyros Trivizas, Pantelis Kalavrytinos

Provincie Limburg
Jef Malliet

Roger Layton Associates
Roger Layton

Royal Museums of Art & History (KMKG)
Roxanne Wyns

Semantika (SEM)
Sašo Zagoranski

SKINsoft ltd. (SKI)
Nel Taurisson

Stiftung Preussischer Kulturbesitz (SPK)
Stefan Rohde-Enslin

System Simulation Ltd (SYS)
Michael Selway, Mike Stapleton

Xantys ltd. (HIM)
Ian Smith

Zetcom AG (ZET)
Norbert Kanter

Appendix 3: Requirement Analysis Task Assignment

1. Definition of the Europeana Connection Kit (ECK)[28]

Suggested definition:

A framework of:

- Software modules;
- Procedural guidelines;
- Documentation;
- Tutorials.

After the framework has been appropriately implemented in a collections management or aggregator's system it will guide users through process of:

- The delivery of standard descriptive metadata to a target (e.g. an aggregator and Europeana) by a content holding institution;
- The possible ingestion of enriched metadata from the target by a content holding institution.

The main types of users of the Europeana Connection Kit are:

- Content providing institutions;
- Software developers;
- Aggregators;
- Europeana.

Your comments:

Please comment even if it is only to agree with the suggested definition.

[28] An example of a toolkit is *Toolkit for the Impact of Digitised Scholarly Resources*. See: http://microsites.oii.ox.ac.uk/tidsr/faqs

2. Assumptions about the ECK

A first list was brought up in our previous survey and/or discussed in Berlin.:

- Software & documentation that is useful and usable;
- Light-touch software;
- Get as much content as possible onto Europeana;
- Straightforward to use (also for people outside the project);
- Should lower financial & technical barriers specifically (but in all barriers in general);
- Speed at all stages;
- Streamlined process;
- Does not take out the human factor;
- Have a preview-possibility (like MINT);
- Transparent and simple;
- No extra costs for content providers;
- Sustainability;
- Show whole workflow;
- Document more than only the ECK;
- Clear roles and responsibilities;
- Basic information about the technicalities;
- Every collection management system should be able to use the ECK;
- Content providers more in control over exchange process;
- ...

Your comments and additional assumptions:

Please at least comment. However we really would like more assumptions from you.

3. Software requirements: Workflow steps

The setup for the Requirement Analysis proposes headings and short descriptions for the conceptual model of the range of functions the ECK is suppose to support. The first parts covers functional software requirements. Please refer to the full description in the lastest version of the *Requirements Analysis Setup* in Basecamp. Revised them, based on your own knowledge and experiences with exchange of cultural content and the systems that support them.

Workflow step:

Manage:
Data management by the collection holders, CMS vendors and aggregators. At the moment it is assumed that there are no additional requirements that should be covered by the ECK.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Select:
Selection of data which will be added to Europeana (or withdrawn for that matter), based on the rights and value of the sources. Europeana Inside will provide granular management tools (part of ECK) that are integrated with CMS and DAM systems that will give collection holders the possibility to include in/exclude from data (records/properties) the Europeana exchange.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Prepare:
Create the data for transfer to Europeana out of the selected data in the CMS.

- ***Metadata cross mapping***: The ECK will provide a drag and drop interface for metadata cross mapping.
- ***Enhancing metadata quality***: The ECK will provide tools which will enhance the quality of the metadata. The data as managed by a content provider takes some knowledge about the collection implicit. Once reused outside of this context the data is multi interpretable at the best. Making implicit knowledge of the collection and it context explicit in the data is necessary.
- ***Apply Persistent Identifier***: Persistent Identifiers (PIDs) are a crucial element in the ECK. But PIDs come in many guises
- ***Apply License***: Making sure that adequate arrangements have been made with regard to owners and usages rights on a granularly base (property, record and/or collection level) and for the cultural object as well as the metadata.
- ***Apply data format***: What data formats are allowed for? XML seems the standard for all CMS involved within Europeana.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Validate:
Validation of the data for ingest in Europeana. Content checker: does the data adhere to the data model of Europeana? A preview can tell whether the data displays well and works well given the intended reuse?

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Exchange (push or pull):
Two way exchange of data between the source and the target. CMS aggregator and Europeana can all be both depending on the context. Deletion is considered a special manifestation of exchange.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Validate:
Formally certify that received data is suited for processing and storage in the new environment.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Accept:
The data can be admitted to the target environment for further processing.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Prepare:
Process the ingested data so it meets all internal requirement of the new environment.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Merge:
The received data should be merged into the collection of data that already exist. This can be considered as a reversed selection.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Additional steps:

Please give any additional steps that are needed, with sub-steps. **If there are no additional steps needed please say so above.**

4. Gathering requirements: User stories

Again based on the experience in your own context you'll be able to describe actual requirements for the ECK either to resolve current problems you're facing or to create optimal enhancements to your current workflow that would make possible or simplify the exchange of cultural content with platforms like Europeana. In most circumstances this will require collaboration between content provider and their technical partners. Do work together where possible.

In order to collect user requirements we would like you to give us brief user stories in the following format:

As a **<role>**, I want **<goal/desire>** so that **<benefit>**

E.g.:

As a **curator in a museum**, I want **to be able to choose individual records, and individual fields of records, for export to Europeana** so that **I can exclude records and fields I do not want to have in Europeana**.

Workflow step:

Manage:
Data management by the collection holders, CMS vendors and aggregators. At the moment it is assumed that there are no additional requirements that should be covered by the ECK.

Your comments and any sub-steps:

Please at least comment. However we really would like more sub-steps if you have them.

Workflow step:

Select:
Selection of data which will be added to Europeana (or withdrawn for that matter), based on the rights and value of the sources. Europeana Inside will provide granular management tools (part of ECK) that are integrated with CMS and DAM systems that will give collection holders the possibility to include in/exclude from data (records/properties) the Europeana exchange.

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

Workflow step:

Prepare:
Create the data for transfer to Europeana out of the selected data in the CMS.

- ***Metadata cross mapping***: The ECK will provide a drag and drop interface for metadata cross mapping.
- ***Enhancing metadata quality***: The ECK will provide tools which will enhance the quality of the metadata. The data as managed by a content provider takes some knowledge about the collection implicit. Once reused outside of this context the data is multi interpretable at the best. Making implicit knowledge of the collection and it context explicit in the data is necessary.
- ***Apply Persistent Identifier***: Persistent Identifiers (PIDs) are a crucial element in the ECK. But PIDs come in many guises
- ***Apply License***: Making sure that adequate arrangements have been made with regard to owners and usages rights on a granularly base (property, record and/or collection level) and for the cultural object as well as the metadata.
- ***Apply data format***: What data formats are allowed for? XML seems the standard for all CMS involved within Europeana.

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

Workflow step:

Validate:
Validation of the data for ingest in Europeana. Content checker: does the data adhere to the data model of Europeana? A preview can tell whether the data displays well and works well given the intended reuse?

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

Workflow step:

Exchange (push or pull):
Two way exchange of data between the source and the target. CMS aggregator and Europeana can all be both depending on the context. Deletion is considered a special manifestation of exchange.

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

Workflow step:

Validate:
Formally certify that received data is suited for processing and storage in the new environment.

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

Workflow step:

Accept:
The data can be admitted to the target environment for further processing.

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

Workflow step:

Prepare:
Process the ingested data so it meets all internal requirement of the new environment.

Your user stories:

Workflow step:

Merge:
The received data should be merged into the collection of data that already exist. This can be considered as a reversed selection.

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

Additional workflow steps

Your user stories:

Please add user stories from as many relevant roles in your organisations as possible.

5. Business requirements

Referring to the latest version of the *Requirements Analysis* setup in Basecamp, you'll also find a part that covers functional business requirements. Again we ask you to suggest under each caption the issues that require attention based on your own knowledge and experiences as heritage institution or software vendor working in the heritage sector.

Requirement:

Strategic planning:
The decision whether to exchange cultural content is a strategic one to be taken by the board of the collection owner. It must bring added value to the institution, be it social or economical.

Your comments:

Please comment, if only to agree.

Requirement:

Quality control:
Exchanging cultural content is of little use if good quality cannot be supported. This brings additional requirements that apply to the working process as a whole

Your comments:

Please comment, if only to agree.

Requirement:

Business model:
All stakeholders participate in the exchange of cultural content because it brings them added value. Business models are a way to discuss and optimise this added value.

Your comments:

Please comment, if only to agree.

Requirement:

Statistics:
Keeping control of the process requires quantitative information.

Your comments:

Please comment, if only to agree.

Requirement:

Operational planning:
The working process of exchanging cultural data needs to be embedded in the organisation. This requires clear definition of tasks, authorities and responsibilities across all stakeholders involved.

Your comments:

Please comment, if only to agree.

Requirement:

Return on investment:
Innovations, like implementing new software and making new partnerships costs money. There has to be a clear idea about the balance between costs and benefits in the long term.

Your comments:

Please comment, if only to agree.

Requirement:

Risk management:
Performing activities bring risks, especially if the activities include interdependencies between autonomous organizations. Assessing the risk and plan for counter management prevents catastrophes.

Your comments:

Please comment, if only to agree.

Requirement:

Business information:
Required to monitor the business process, report on it and adjust it when necessary.

Your comments:

Please comment, if only to agree.

Requirement:

Contract management:
A way to control the risks when working with partners is contracting.

Your comments:

Please comment, if only to agree.

Requirement:

Data modelling:
A clear understanding how cultural content is collected and managed within the own organisational context is required in order to think about how to make crosswalks into another context of use, like Europeana.

Your comments:

Please comment, if only to agree.

Requirement:

Partner management:
Working together often requires long term interdependencies build on trust. How to find the right partner, make clear agreements and keep the relation sound.

Your comments:

Please comment, if only to agree.

6. Assigned tasks

In the previous survey reference to pre-existing sources was questioned. Based on the answers we divided partners up in teams. Each team is asked to use the different Europeana project or aggregation context listed below as source of information. Within these contexts documentation will exist that contains knowledge that Europeana Inside should reuse or build upon. Please add this knowledge to the collaboratory by bookmarking en annotating it through the Diigo service (instructions document in Basecamp will follow shortly). Alternatively, if it's not online, upload and annotate (part of) the document onto Basecamp.

- ATHENA;
- Linked Heritage;
- HOPE;
- Europeana Local;
- DCA;
- BHL-Europe;
- Culture Grid;
- Culture.fr;
- ErfgoedPlus.be.

Project or working context	**Partners to report on this information source**	
	Content Provider	**Technical Partner**
Athena	FAB, BEN	MON, ADLIB
Linked heritage	SPK	ZET
HOPE	PIM	KE
Europeana Local	SLV	SEM
DCA	NAG	PS
BHL-Europe	RBINS	SYS
Aggregators	MNN	HIM, SKI, KUL
Minimum scenario	SEI	MOB

7. Time schedule

- August 2-3: Distributing work plan;
- August 6-10: (Desk) research;
- August 13-17: Collecting input;
- August 20-24: Draft version of requirements analysis;
- August 27-31: Final editing;

August 31, 2012: Delivery to the Commission.