Grant Agreement 297292

EUROPEANA INSIDE

Minutes of 1st Technical Partners meeting

Document number	*D1.5*
Dissemination level	*Public*
Delivery date	*November 2012*
Status	*Final*
Author(s)	*Carolien Fokke (Collections Trust)* *Isabell Ehrlicher (SPK)* *Monika Hagedorn-Saupe (SPK)* *Stefan Rohde-Enslin (SPK)*

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

Revision History

Revision	Date	Author	Organisation	Description
v0.1	30-10-2012	Carolien Fokke	CT	Draft Version 1
v0.2	01-11-2012	Isabell Ehrlicher	SPK	Draft Version 2
v0.3	14-11-2012	Monika Hagedorn-Saupe, Stefan Rohde-Enslin, Isabell Ehrlicher	SPK	Draft Version 3
v0.9	21-11-2012	Carolien Fokke	CT	Review of draft
v1.0	06-12-2012	Carolien Fokke	CT	Incorporation of participants' comments and production of final version

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Minutes 1st Technical Partners meeting

Place and Date

The Hague, 26th of October 2012

Venue

DEN Offices

Background to the meeting

This meeting is the 1st standalone Technical Partners meeting for the project. Its aims are to:

- Discuss the content partners' requirements;
- Discuss the deliverables from WP 2, their inputs and their outcomes;
- Discuss the functional requirements for the Europeana Connection Kit (ECK);
- Discuss the technical specifications for the ECK;
- Discuss the possible architecture of the ECK;
- Discuss the next steps within the project.

Participants

Name	Job Title	Organisation	Country	Email
Sam Alloing	Business Consultant	Catholic University of Leuven	BE	sam.alloing@libis.kuleuven.be
Dimitra Atsidis	Jr. Data Ingestion Specialist	Europeana Foundation	NL	Dimitra.atsidis@kb.nl
Ágoston Berger	Developer	Monguz	HU	aberger@monguz.hu
Vincent Bodinier	Technical Consultant	KE Software Ltd	UK	Vincent.bodinier@man.kesoftware.com
Eric de Cacqueray	Director	Mobydoc	FR	eric.de-cacqueray@mobydoc.fr
Sam Coppens	Project Engineer	Interdisciplinary Institute for Broadband Technology	BE	samcoppens.werk@gmail.com
Zoltan Csaki	Librarian	Petofi Literary Museum	HU	csakiz@pim.hu
Jean-Yves Cueille	Developer	Mobydoc	FR	Jean-Yves.Cueille@mobydoc.fr
Bert Degenhart Drenth	General Manager	Adlib Information Systems	NL	bert@nl.adlibsoft.com
Isabell Ehrlicher	Research Assistant	Stiftung Preussischer Kulturbesitz	DE	I.Ehrlicher@smb.spk-berlin.de
Berry Feith	Project Officer International Activities	Digitaal Erfgoed Nederland	NL	berry.feith@den.nl
Alex Fell	Technical Operations	KE Software	UK	alex.fell@kesoftware.com
Carolien Fokke	European Projects Officer	Collections Trust	UK	carolien@collectionstrust.org.uk
Monika Hagedorn-Saupe	Deputy Director	Stiftung Preussischer Kulturbesitz	DE	m.hagedorn@smb.spk-berlin.de
Christ Hagenaars	Manager Software Development	Adlib Information Systems	NL	c.hagenaars@adlibsoft.nl
Arthur Hanselman	Project Manager and Consultant	Gallery Systems & CIT	NL	arthur@go2cit.nl
Wietske van den Heuvel	Web Editor	Digitaal Erfgoed Nederland	NL	Wietske.vandenheuvel@den.nl
Antoine Isaac	Scientific Coordinator	Europeana	NL	antoine.isaac@kb.nl
Mark Johnson	Developer	Knowledge Integration	UK	Mark.johnson@k-int.co.uk

Name	Job Title	Organisation	Country	Email
Norbert Kanter	Managing Director	Zetcom	DE	Norbert.kanter@zetcom.com
Jörg Kruschinski	Manager	Zetcom	DE	Joerg.kruschinski@zetcom.com
Monika Lechner	Quality Manager Digital Heritage	Digitaal Erfgoed Nederland	NL	Monika.lechner@den.nl
Vincent Leconte	Developer	Mobydoc	FR	Vincent.leconte@mobydoc.fr
Jef Malliet	Ingenieur	Provincie Limburg	BE	jmalliet@limburg.be
Yorgos Mamakis	Back-Office Tool Developer	Europeana	NL	Yorgos.mamakis@kb.nl
Zoltán Mazula	Developer	Monguz	HU	zmazula@monguz.hu
Gordon McKenna	International Development Manager	Collections Trust	UK	gordon@collectionstrust.org.uk
István Nagy	Developer	Monguz	HU	inagy@monguz.hu
Jo Rademakers	Head of LIBIS	Catholic University of Leuven	BE	Johan.rademakers@libis.kuleuven.be
Stefan Rohde-Enslin	Scientist	Stiftung Preussischer Kulturbesitz	DE	s.rohde-enslin@smb.spk-berlin.de
Michael Selway	Managing Director	System Simulation Ltd	UK	mas@ssl.co.uk
Neil Smith	Director	Knowledge Integration Ltd	UK	neil.smith@k-int.com
Marco Streefkerk	Senior Consultant ICT & Cultural Heritage	Digitaal Erfgoed Nederland	NL	marco.streefkerk@den.nl
Nel Taurisson	Research and Development Engineer	Skinsoft	FR	Nel.taurisson@skin-soft.org
Rob Tice	Director	Knowledge Integration	UK	Rob.tice@k-int.com
Spiros Trivizas	Senior Developer/Consultant	PostScriptum	GR	strivizas@postscriptum.gr
Roxanne Wyns	Project leader ICT - Applications	Koninklijke Musea voor Kunst en Geschiedenis	BE	r.wyns@kmkg-mrah.be
Sašo Zagoranski	CTO	Semantika	SL	Saso.zagoranski@semantika.si

Agenda
26th of October

Agenda item	Minute
Welcome and Tour of the Table	Gordon McKenna (GM) from Collections Trust gave a short introduction to the meeting and all participants introduced themselves shortly.
Video about Europeana Inside (CT)	A short video by Nick Poole, Collections Trust CEO, was shown and a discussion of the project vision presented in the video followed. Link to video: http://www.youtube.com/watch?v=9Fw3EPJ7tNc. GM asked the Management Board to share their ideas about the project's vision. Neil Smith (NS) from Knowledge Integration expressed his view that the name of Europeana Inside is supposed to refer to 'Intel Inside'. The goal of the project is to seamlessly connect any collections management system (CMS) to Europeana. This is why there are so many different technical partners (TPs) involved. NS joined Nick Poole in the project's vision that the software will stand for having 'Europeana inside.' Marco Streefkerk (MS) from Digital Erfgoed Nederland shared the view to make collections available for Europeana and other portals without complicated technical handling. He said it is crucial to concentrate on how to prepare knowledge gathered during the project for reuse. It is important to make Europeana content available for creative and commercial reuse etc. Roxanne Wyns (RW) from Royal Museums of Art and History in Brussels said that it is important that data are presented in an interoperable format, for Europeana but also for other portals and projects on national level. The Europeana Connection Kit (ECK) will allow cultural heritage institutions' collections to get a better connection to the outside. It is important that the ECK should be developed for easy and simple use to save working time. Monika Hagedorn-Saupe (MHS) from Stiftung Preussischer Kulturbesitz said that as a museum representative she believes museums need help now to share the content in their CMSs. She would like to have more museum data available on Europeana. She mentioned that many systems are unable to export data easily. The technical support and tools for institutions data export to make data available is crucial. GM invited the TPs to share their vision of the project.

	Jörg Kruschinski (JK) from Zetcom saw as an outcome simplifying harvesting and mapping processes.Norbert Kanter (NK) from Zetcom said that the outcome should be that the process of adding content to Europeana should be easier than it is at present, also for small and medium museums. Nevertheless the ECK cannot respond to everything. Bert Degenhart Drenth (BDD) from Adlib said that it should be as easy as using iPhoto which means that most of the functions work in the background. Jo Rademakers (JR) from the University of Leuven stressed that also archives and libraries should be involved in the process. GM mentioned that there are financial barriers preventing institutions to take part. He also pointed out the continuing costs of software vendors. The Europeana Inside project and the funding of the Commission supports overcoming these financial barriers in developing a tool. The ECK should not be available on extra costs but it should be in the software price. There cannot be an additional charge for the ECK, but perhaps there could be for the services. To make sure that this happens it is important to stick to standards. In addition the ECK source code should be open source for free re-use. Jef Malliet (JM) from Provincie Limburg stated that it is also a question of trust. People should understand what is happening to their content. He believes that this is also a job of the project as a whole. He believes the Europeana Inside project should go a little further than only producing software tools. GM mentioned in response to JM's remarks that communication is key. Therefore he stated that as a first step Collections Trust (CT) was going to ask content providers (CPs) to look at the DOW and check whether there are any terms in there they did not understand. He also invited the TPs to look at the DOW and imagine what they might want to know if they were a CP. Marco de Niet (MdN) from Digital Erfgoed Nederland said that he was surprised to hear the statements about Europeana in the video. He thought that Europeana were widening their approach. MdN stressed the role of the Europeana Network. Monika Hagedorn-Saupe (MHS) responded that on the 27th of November in Berlin the annual Europeana Network meeting will take place. There, a discussion about Europeana's wider focus will take place. Furthermore, in the future Europeana wants to focus on:

	• Strengthening the network; • Being a facilitator, allowing technical re-use; • Creating wider API-possibilities. MdN also mentioned that many institutions do not understand the strategies. He also referred to 'DE BASIS' (The Basics), a concept DEN has developed about interoperability. It focuses on Basic Practice and not on Best Practice (as this is sometimes not attainable for smaller institutions). DE BASIS will be translated into English in the context of Europeana Inside by DEN and CT together. **ACTION: CT to ask Content Providers what terminology from the DOW they do not understand, and put together a glossary.**
Presentation content providers requirements (KMKG)	RW gave a presentation about the requirements discussed by the CPs during the additional Content Providers meeting in Brussels on the 8th of October 2012. She discussed the requirements per workflow step (the workflow steps were developed by DEN during *D2.1: Requirements Analysis*). <u>Step 1: Manage</u> BDD said it seemed to him as if the process consisted of 2 steps, and it was not a 1-click button. RW explained that exporting metadata is often a problem. BDD mentioned iPhoto again; using it to discuss the idea that while there are many steps to publishing a photo, they are all invisible and it seems as if it is all just one click on a button. He even referred to the introduction video and the drawing of a box on the website which implies that things are invisible. RW replied that iPhoto only works for one target. NK said that he did not think that a) export, b) edit, and c) send exported and edited metadata are part of the ECK. He said this should all be in the CMS. RW said she agreed, but she used the example of editing 30.000 records. She said there is no possibility to do this in a batch, and that she has to open each and every one of those records to edit them. Because Europeana requires quite specific information that is quite useless in the CMS itself, it is necessary to add this information once it has been decided to send the records to Europeana. This is very time-consuming. NK then said that batch processing should be part of the ECK, but all the rest is part of the CMS. MS tried to clarify the requirements further. He explained that usually the CMS is purely for internal use. It is different once it comes to exchanging information, because then you have to add target-specific information to your records. Most CMSs have modules for making exhibitions etc., but those are different modules.

	Ágoston Berger (AB) from Monguz then explained that several institutions do not have a proper CMS and only have a simple Excel file as a CMS, and not an official CMS. There should be a tool to do this process. RW then reiterated that she cannot do batch processing and that others have also expressed that they cannot do this but that they would like to be able to. Michael Selway (MSL) from System Simulation said that these are more like mapping issues. He said that it would be possible to set up a mapping saying something like: 'we cannot set up because of the following kind of reasons' or 'we do not have rights to distribute the photograph'. He stressed that the ECK and the CMS should not be separate, but that the ECK would be part of the CMS. 'The ECK is not an App', it is lots of software that software providers may want to implement or may not want to implement. There is no need to export to a file. NS said that everyone needs to use their person months in the project to ensure that the ECK is embedded in the CMS, but that the ECK should also be an implementation for people who do not use a CMS. MSL then said that he does not want to create a competing product and BDD agreed that the purpose is not to build software that works without a CMS. NS then said that it could be discussed later in the afternoon. BDD added that the project is called Europeana Inside pointing to the box on the website that something is inside of something else, and should work as simply as possible. RW pointed out that the export is not mentioned in the ECK description. She hoped to discuss this in the afternoon. If provider information needs to be added, it should be managed in the ECK. GM added that it could be done in the CMS and in the process, mentions that it is like a translation. NS mentioned the technical specification of the ECK to embed the tools in the technical partners' software and to make it transparent for the majority of cultural heritage institutions using CMSs. MSL stressed that it is important for technical partners not to be in competition. BDD added that producing products allowing sending data from Excel files to Europeana is possible. RW concluded that the ECK should deal with all kind of formats.

	<u>Step 2: Select</u> Alex Fell (AF) from KE Software wanted to know why there was a requirement to upload only new records. RW explained that this was because some of the records would be updated and others would not be. AF then suggested thinking of that in terms of mapping automatically, but RW said that curators want to be in control of updates to records etc. MSL said that this sounded like Europeana is an archive, more than anything. RW said that it might be fine to send people a warning about a possible update but that making it entirely automatic was not a good idea. Once objects are on Europeana for the first time, it might be ok to update (with a warning) after that. RW mentioned that full control of export is necessary e.g. harvesting by Europeana should be a choice. Institutions have to document what was sent to Europeana. Curators are involved in discussions with directors etc., which is why a control of export is absolutely necessary. Someone from Europeana proposed to send it as a batch or as a continuous update and asked whether automatic updates are necessary. RW confirmed that an automatic update is needed. MSL mentioned that this is new for him. RW added that the update of the first batch is important. Updates could be very beneficial after the batch has been sent. BDD stressed that it is important not to mix the terms 'export' and 'published information'. He proposed to differentiate between those terms. NK and RW proposed that messages like 'agreement on update' or a warning like 'you can still disagree not to update' can be used. NS concluded that lots of content providers do not want to have an automatic update. <u>Step 3: Prepare</u> RW said that PIDs need to be present if there is any chance of doing automatic updates (with a warning). In the step 'Prepare' the target is selected, mostly data is first published on national portals and then on Europeana. It should be possible with batch processing to add constant values to collections e.g. 'Africa' to an African collection.

	Step 4: Validate RW described the requirements for Step 4: Validate. For mandatory elements you should get an error if information is missing. Step 5: Supply (push/pull) AF brought up the discussion about the role of the aggregators again: smaller institutions may not have the possibility to use OAI-PMH. Not all aggregators use OAI. OAI can be used to implement harvest. NS explained that at the moment this project was suggested, Europeana was not going to accept content directly from CPs. Maybe Europeana can be persuaded in the future to work with the content providers if the ECK is available. More direct supply to Europeana depends on the success of the project. Therefore, the architecture should now cope with both direct supply and supply via aggregators. MSL said that it would be a similar system for aggregators as for Europeana. RW said that if the tool is simple and easy to use, content providers can supply on their own. She does not know many aggregators who use the ESE/EDM format as standard though. Step 6: Data Acceptance RW explained the requirements for Step 6: Data Acceptance. There were no comments. Step 7: Enrich & Return RW explained the requirements for Step 7: Enrich & Return and said that a control of the flow back is necessary e.g. geo-names, UGC. There were no comments.
Scene Setting: Process for producing *D2.5: Technical Specification* and *D3.1: ECK Iteration 1* (K-INT)	NS gave a presentation about the process and schedule for *D2.5: Technical Specification* in November 2012. When WP 3 and 5 were examined, it became clear that the time schedule did not fit with WP 4. There was a suggestion to adjust deliverable dates. NS explained the initial idea had been a 36-month project with a waterfall approach. Already during negotiation meetings with the Commission it became clear that if the project was shortened, a more iterative approach was needed. Part of this new approach was implemented in the Description of Work (DOW). D2.4 will be refined throughout the project, there is time to change it in later versions. In WP 3 and WP 5 no additional deliverables have been planned, nor later delivery dates. However, for clarification some of the deliverable names were changed a bit. NS presented the new iterative development plan with an iterative approach commonly used in the technical field.

	NS then discussed what each iteration should contain Iteration 1: Must-functionalities; Iteration 2: Management functionalities; Iteration 3: Re-ingestion functionalities; Iteration 4: Everything. Iteration 1 is to be finished in 5 months. It should include everything which will be needed, also the step 'Select'. Iteration 2 runs until month 18. The developments start in month 15. If something does not work, it can be taken into the next iteration. The 2nd iteration should include what Europeana expects. Iteration 3 should include the minimum and the ingestion functionality (WP 4 requirement). Iteration 4 includes everything that can be done. WP 5 starts in month 21. NS mentioned that *D3.2 Codebase* can perhaps be posted on GitHub (to be discussed in the afternoon). *S5.0* is not mentioned in DOW, but addresses the content ingestion process which should not be at the end of the project. The meetings fit well into the plan; deliverables should be connected to the meetings. NS ended the presentation of the development plan by saying that developing software from November 2012 to July 2014 and working on different iterations is possible. He announced the presentation of the proposal of the Technical Architecture from Rob Tice (RT) which would be in Iteration 1. NS discussed that some of the dates for the TP meetings that had been scheduled before now seem to not coincide with the development of the various iterations. He suggested talking about new dates, but this did not happen because it was still too early in the software development process. MS asked whether some time should be reserved in the schedule for implementing the new software. NS answered that he suspected people would implement it in the most recent version of their software. However, the implementation does need to be aligned with WP 4 and the customers' needs. BDD said that the 1st priority is to make the software work, and the 2nd priority is to plan how to implement it.

	GM asked about *s2.6* and *s2.7* in the new schedule: who will deliver these? NS said they would deliver these. He also mentioned that the TPs should think about how the new requirements that might surface during testing should be incorporated in later iterations. NS responded that this was difficult to say as it is still unclear how much access CPs will have to the ECK software in iteration 1. RW said that if you read the DOW carefully it does not state that the content has to go into Europeana straight away. Antoine Isaac (AI) from Europeana said that it is important that when thinking about the timing, there is communication with Europeana as well. Because some of the requirements are related to Europeana, they have to be aware of what is needed and/or expected as well. BDD then suggested a formal contract about the cooperation between the TPs and Europeana and what Europeana would offer and make possible. This was not discussed further. GM mentioned that Europeana might be part of the Management Board, or at least take part at the Management Board meetings. **ACTION: CT to contact Europeana about being part of the Management Board.**
Proposed Technical Architecture of ECK (K-INT) & Mapping Requirements to the proposed Architecture (K-INT)	Rob Tice (RT) from Knowledge Integration presented a proposal for the architecture of the ECK. He presented a possible approach to what the ECK will be. He said that the ECK will be a set of loosely coupled components which can be deployed locally, remotely, collectively or separately to fulfil all or part of the ECK functional requirements. He described broadly that there is an input, an output, and in between there is the ECK. He mentioned that the ECK is made of some components marked in several colours on the image he presented. The input is the user choosing which records go into Europeana. The output is the records that are published on Europeana. He stressed that he was not trying to be prescriptive, and that he was aware that some functionalities and requirements are closer to the CMS than to the ECK in a way. The reason for starting out so broadly is that there are so many people involved. RT suggested that when talking about manage and select, it is difficult for the ECK to be prescriptive because each CMS is different. MSL said that the requirement is wrong in that case.

	In the case of a CP without a CMS who wants to use the ECK there is the need for reference implementation. MSL questioned whether this is in the project's scope. RT explained that this was not the case; it was simply providing a service for people who only have Excel files with their records, for example. NS added that this could also refer to aggregators. BDD also thought this is outside of the project's scope. MSL said that the DOW says it should be open source. RT added that the open source will not be controversial. RW added that the DOW also states that it should be possible to go into Europeana without using an aggregator. RT said that every tool should be available for people with Excel-based systems. MSL disagreed. NK added that this function does not have to be developed. An institution working with spreadsheets will deliver to aggregators. Sašo Zagoranski (SZ) from Semantika mentioned that the mapping functionalities should be named and explained to see which of the steps can be made easier. BDD pointed out the general requirement 'make publication to Europeana easier' in the DOW. This basically means that there is a requirement for Europeana to be aware whether there will be both the possibility for CPs to deliver directly to Europeana or to go through an aggregator. MSL mentioned that Europeana should be part of all discussions to get to know which interface will be developed or which interface should be developed. BDD mentioned that the requirements are not listed yet. It is necessary to give the list to Europeana to get feedback. BDD added that the next step should be to talk with the implementers how to treat the requirements on a functional level and then to define on the backend how to reach this solution. Everyone is ok with testing whether CMSs are ECK-compliant and whether aggregators are ECK-compliant. RT said that a good starting point will be to incorporate the musts from *D2.4* first as implementations in iteration 1. The Europeana API should not yet be considered for iteration 1. **ACTION: NS to send wish list to Europeana about what we need to know and what will be the future for Europeana.**
Group working on specific functional areas	See the minutes of the working groups in the Appendix.
Feedback from groups (K-INT)	After the working groups about the functional requirements of the ECK and the identification of the 'candidates' for Iteration 1, the results were shortly presented.

	<u>Group 1 Functional Areas 'Manage' and 'Select':</u> Mark Johnson (MJ) said that the management export is often implemented in the CMS, but reporting from the ECK and storing back in the CMS could be possible. Event modification should also be sent back to the CMS from the ECK. The selection-possibilities are provided by the CMS itself. Requirement 2.07 'Reuse saved queries' is very popular, and should be a 'MUST'. All 'MUST' requirements will be provided for iteration 1. <u>Group 2 Functional Areas 'Prepare' and 'Validate':</u> RT said the ECK could be built on LIDO. In iteration 1 there would be an export to LIDO and select mappings to LIDO in the CMS. Some questions remain about the requirements. 3.10 'Multiple Assets' and 3.11 'Defining media types' are both 'MUST' requirements. <u>Group 3 Functional Areas 'Supply' and 'Acceptance':</u> Carolien Fokke (CF) from Collections Trust summarised that the participants in group 3 were of the opinion that WFR.05.01 'Automatic supply' actually consisted of two requirements: push and pull. There might be a pilot for data push back to Europeana. WFR.05.02 'RE-supply functionality for failed records' and WFR.06.01 'Preview presentation Europeana' need more clarification.
Wrap-up and forward plan (K-INT)	NS discussed the final actions to be taken. RW stressed that it is important to inform the CPs of the process and give them a bit of education in the use of standards etc. The TPs might be able to support this. GM stated that the set of actions is clear and reminded everybody to send deliverables in time. **ACTION: all groups to supply notes to Carolien Fokke (CF) and Isabell Ehrlicher (IE) for the meeting minutes document.** **ACTION: NS to make a list with questions for EUR (also about EDM).** **ACTION: A good draft of *D2.5* should be ready by the end of the 2nd week of November.** **ACTION: BDD to send ideas about architecture next week to NS who will then circulate it.** **ACTION: TPs to contact their CPs directly with questions about requirements and functionalities. MS / DEN will help putting TPs into contact with specific CPs who supplied requirements and functionalities.**

APPENDIX I: GROUP WORK

Minutes Group 1: Manage, Select

Participants: Mark Johnson (leader), Berry Feith (notes), Ágoston Berger, Alex Fell, Bert Degenhart Drenth, Gordon McKenna, Jean-Yves Cueille, Jörg Kruschinski, Monika Hagedorn-Saupe, Monika Lechner, Roxanne Wyns.

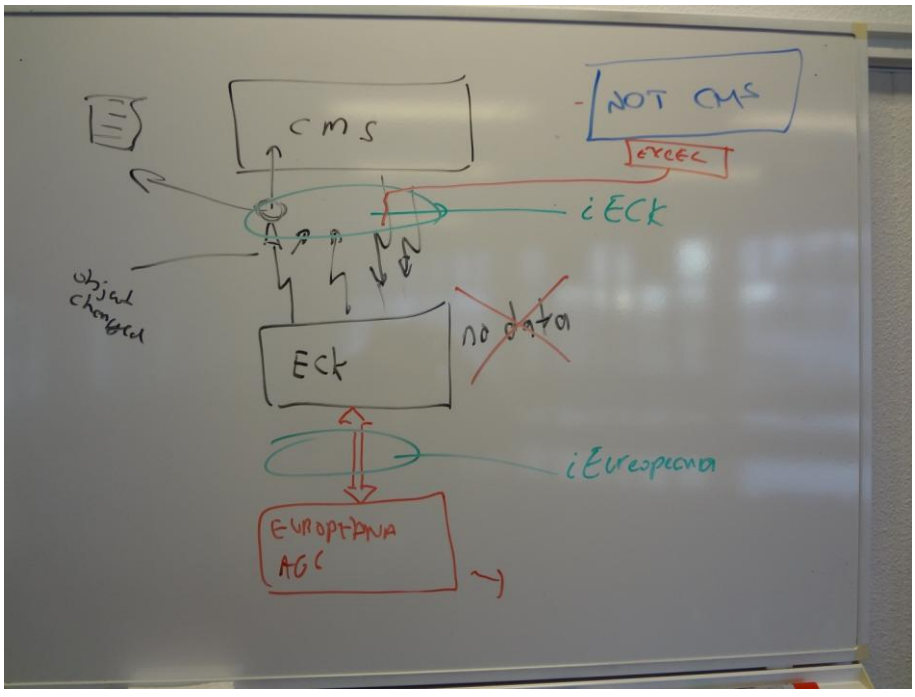
It was discussed that it is difficult for the ECK to be prescriptive about how to manage and select data as there are so many software vendors. The Europeana API communicates the PIDs Europeana added and offers to get enriched data back into the ECK.

Roxanne Wyns (RW) stated that identifiers for collections and institutions are needed, too. Bert Degenhart Drenth (BDD) said that there is a lack of knowledge on how PIDs are allocated or used. Monika Lechner (ML) commented that the ECK connects data to Europeana. It is not the CMS that is connected to Europeana.

It was told that there are CMSs that cannot produce PIDs. Jörg Kruschinski (JK) asked where exactly the relevant PIDs are produced – in the CMS or ECK – and added that some parts of the log files might be in the CMS.

ML commented that the ECK will be Open Source and that the aim should be to put as much as possible into the ECK. BDD stated that the ECK has to be a library of building blocks which everybody can implement in their CMS. ML pointed out the necessity to add an 'update sign' and to store this information in the CMS to make it traceable in the future. Gordon McKenna (GM) asked where the information will be if the ECK disappears. RW proposed that the log files of the ECK should flow back into the CMSs. BDD added the ECK should produce event logs which CMS vendors can choose to integrate in the CMS or not.

BDD drew his vision of the ECK on the whiteboard:



Picture: Monika Hagedorn-Saupe (SPK)

Two interfaces: iEuropeana

iECK

ML asked for the meaning of 'iEuropeana' and whether it was meant as an API. Ágoston Berger (AB) mentioned that the ECK should consider a lot of protocols. BDD stressed that it is necessary to know how Europeana handles things now. Europeana does not get data automatically.

BDD explained his drawing further and mentioned that the two interfaces were visualized to have a framework to work from:

- no single architecture
- simple components (API is a reference to OS implementation)
- functionality / components / services are separate entities

BDD said that TPs need to build reusable components. The communication and functionality of different components of the ECK have to be defined. Different versions or components of the ECK are:

1- CMS integrated ECK
2- ECK local
3- ECK remote
4- If no CMS is used, ECK local and/or ECK remote

This led to the question whether all these 3 (or 4) have different functional requirements.

It was said that the ECK is for aggregators and content providers alike.

BDD mentioned that the technical team needs to examine the way things are done currently. It should build on what is already there. Europeana should be contacted as well.

Manage – Requirements Overview

This workflow step describes all aspects of data management and user management.

No.	Requirement	Explanation	Priority	ECK or CMS?
WFR.01.01	Export management	The system is able to tell which records have been exported when to Europeana.	Must	**CMS** But event log from the ECK as feedback for the CMS.
WFR.01.02	Revision history	The system is able to show which records are altered when and by whom, so it can provide a base for updating exported records.	Must	**CMS** Feedback of success.

~~WFR.01.03~~ HLR	Notification changes to the ECK	The system transmits a notification when changes are made to the ECK that might have an impact on the local management. Feedback on (system) updates.	Must	***High Level Requirement**
WFR.01.04	PID management	The system manages PIDs for objects that can be used for identification when data is sent to Europeana.	Must	**CMS** If Europeana builds one: event log.
WFR.01.05	Enriched data management	The system is able to merge and manage returned enriched data once ingested in the system of the CP.	~~Could~~ Must	**

***The meaning of the marked requirement 'High Level Requirement' was asked to be explained. GM answered that EDM may change. Are there implications that such a change brings changes to the CMS? In any case it has to be a high level requirement.

**It was said that WFR.01.05 Enriched data Management can be done after Iteration 3 and not at the beginning. It has to be tested as it is mentioned in the DOW. It was first marked as a 'Could' but should become a 'Must' if it was successfully tested.

It was stressed that the ECK is not a repository.

The discussion went back to BDD'S drawing. Should the iECK have a user interface or not? There was an understanding that the iECK must have a socket that can integrate records for CPs and Aggregators which want to provide cross collection and cross CMS data.

There are multiple options at this step including a number of default mappings.

Select – Requirements Overview

This workflow step describes the selection process. Everybody agreed that all of this is going to take place in the CMS, so these are not functional requirements for the ECK.

No.	**Functionality**	**Explanation**	**Priority**	**ECK or CMS?**
WFR.02.01	Selecting multiple records	The system can make a selection of multiple records.	Must	**CMS***
WFR.02.02	Selecting a single record	The system supports making a manual selection of multiple records or a single record.	Must	**CMS***
WFR.02.03	Selecting records based on values	The system is able to select records based on specific values in a variety of fields: e.g. by location, by object category, by theme, by section, or by (part of) inventory number.	Must	**CMS***

WFR.02.04	Boolean operators	The system is able to combine filters with clear Boolean operators.	Must	**CMS***
WFR.02.05	Indication of selected fields	The system shows whether certain records or fields are or will be included in a selection.	Must	**CMS***
WFR.02.06	Selecting within records	The system is able to exclude or include each individual digital asset attached to a record in a selection.	Won't	**CMS***
WFR.02.07	Reuse saved queries	The system is able to repeat a certain selection, e.g. for updates, so filters or queries must be storable and re-usable.	~~Could~~ Must (this is one of the most popular functions of a CMS)	**CMS***
WFR.02.08	Managing multiple selections	The system is able to manage multiple selections, for Europeana and for various aggregators. Selections can be based on different criteria and can overlap.	Won't	**CMS***
WFR.02.09	Standardized selection filters	The system is able to exchange selection settings (filters, criteria, etc.) through the use of a uniform XML schema.	Won't	**CMS***

*Everybody agreed all should be part of CMS. The requirements have to be fulfilled in the CMS to work with the ECK.

Minutes Group 2: Prepare, Validate

Participants: Rob Tice (leader), Isabell Ehrlicher (notes), Christ Hagenaars, Dimitra Atsidis, Eric de Cacqueray, István Nagy, Jef Malliet, Jo Rademakers, Nel Taurisson, Norbert Kanter, Vincent Bodinier, Wietske van den Heuvel

Rob Tice (RT) welcomed everybody and introduced the task of working group 2: to discuss the musts of the functions in the steps 'Prepare' and 'Validate'. The group started with the step 'Prepare' which contains all activities related to data preparation.

At the beginning Norbert Kanter (NK) from Zetcom asked whether there is more than one mapping. In his eyes there should be one transformation file. It was mentioned that some collections are well documented and that different mappings exist. The first mapping will lead to an intermediate one. NK asked where the position of the input format is. Jef Malliet (JM) from Provincie Limburg asked for a definition of 'predefined input'. RT mentioned that the definition of 'predefined input' will be clarified. NK added that the system needs to transform metadata from predefined LIDO to EDM. NK thinks that the ECK should do an automatic transformation. Eric de Cacqueray (EdC) from Mobydoc wanted to know why LIDO is proposed and pointed out that maybe some do not know LIDO. JM added that maybe there are many systems in use. It was commonly understood in the working group that the intermediate format has to be clarified. JM added that it will be easier to map to LIDO or MARC and then to EDM. RT mentioned that the EDM format is very specialised. JM remarked that the ECK should be outside the CMS.

EdC wanted to have a more precise definition of 'configuration during the system'. JM added that a distinction between system and ECK is also necessary. The group mentioned that LIDO, MARC and others can be used to convert metadata to EDM. If LIDO is used for the ECK a definition of a certain LIDO profile is necessary for Europeana, the same has to be considered for MARC. Soon a discussion with the content providers and Europeana will be necessary.

After this introductory discussion, the working group started commenting on the functionalities of the step 'Prepare' in Iteration 1 and began to define the 'musts'.

Prepare

WFR.03.01 Automatic EDM mapping

There should be an intermediate 'transport' format per institution type and the ECK should only deal with managing and validating this (including image data). LIDO was proposed for museums (EAD and Marc for Archive and Library domains). It was understood that a Europeana Inside profile of LIDO would need to be defined. The selectable mappings should be from the CMS internal format to this LIDO profile. It should also be possible for CMS vendors to choose to use EDM directly. If the CMS is not LIDO compliant, the ECK should be able to be extended by adding plugins to deal with other formats. There should be no editing of the mapping from LIDO to EDM. There is no export by now from a CMS to Europeana, it always goes via an aggregator.

WFR.03.02 Preview mapping

In this functionality the working group came up with the necessity of having the possibility to prove the quality of the converted metadata including a thumbnail.

WFR.03.03 Editable mapping

In this functionality the mapping can be edited to correct/improve the metadata conversion from source to target data model. The CMS to LIDO-mapping was mentioned. It was stressed within the working group that this step is essential to test with content providers at early stage.

WFR.03.04 Mapping feedback

JM considered the mapping feedback as a must for Iteration 1. JM told that sometimes transformation stops and feedback is required to solve the problem. A short message about the problem will do.

WFR.03.05 Saving mapping

The working group considered this functionality as a must for Iteration 1. There were no more comments.

WFR.03.06 Field explanation:

The working group considered this functionality as a must for Iteration 1. There were no more comments.

WFR.03.07 Automatic value insertion

The working group considered this functionality as a must for Iteration 1. There were no more comments.

WFR.03.08 Automatic thumbnail generation

As it was discussed in functionality ‚WFR.03.02 Preview mapping' the automatic thumbnail generation is a must functionality to be developed in Iteration 1. One part of this functionality is the validation, i.e. to test whether an image is available. A second part is the thumbnail generation; however, Europeana generates its own thumbnails so that this generation might just apply to the previewing.

WFR.03.09 Thumbnail selection

The working group considered the identification of assets as part of CMS functionality so image selection and use is part of the mapping step. It was discussed that the first thumbnail can be chosen in the functionality 'Thumbnail selection' which depends on the matter of the mapping. JM told that the first image is used as main thumbnail, most of the people will do.

WFR.03.10 Multiple assets

The working group considered this functionality as a must for Iteration 1 and should be changed to a must.

WFR.03.11 Defining media types

The working group considered this functionality as a must for Iteration 1 as many collections are mixed collections using different types. As the EDM media type is mandatory this function is a must. The LIDO profile needs to use the EDM media types.

WFR.03.12 Metadata field on IPR digital object

The working group considered this functionality as a must for Iteration 1.It is part of the mapping functionality but might be accessed and configured via a specific user interface.

WFR.03.13 Metadata field on IPR metadata

The working group considered this functionality as a must for Iteration 1. It is part of the mapping functionality but might be accessed and configured via a specific user interface.

The discussion stopped here as the time ran out. RT added his own comments to the other functionalities in step 'Prepare' to the document about the functionalities of each step after the meeting: WFR.03.14 Metadata field on IPR preview, WFR.03.15 Mark mandatory fields, WFR.03.16 Choosing a default mapping and WFR.03.17 Automatic data suggestion.

RT added his own comments to the functionalities in step 'Validate' to the document about the functionalities of each step after the meeting.

Prepare – Requirements Overview

No.	Requirement	Explanation	Priority
WFR.03.01	Automatic EDM mapping	The system converts metadata automatically from a predefined input format to EDM by (a set of) default mappings that is selected during configuration of the system.	Must
WFR.03.02	Preview mapping	The ECK shows a preview of the converted metadata and associated thumbnails that are the result of applying a specific mapping. It also indicates the quality of the converted metadata including the thumbnail.	Must
WFR.03.03	Editable mapping	The mapping can be edited to correct/improve the metadata conversion from source to target data model.	Must
WFR.03.04	Mapping feedback	The system reports on problems with applying the mapping.	Must
WFR.03.05	Saving mapping	The system saves the mapping for repeated use.	Must
WFR.03.06	Field explanations	The system informs on the expected input required for the concerned fields in the mapping.	Must
WFR.03.07	Automatic value insertion	The system is able to insert constant values automatically for metadata not included in the collection database as defined by the user, e.g. language of record, content provider name.	Must
WFR.03.08	Check digital asset availability	The system ensures that an image is available for access by Europeana or other targets to generate a thumbnail.	Must

WFR.03.09	Thumbnail selection	If more than one digital asset is linked to a metadata record the system can choose which image will be used to produce a thumbnail based on input of the user manually or in batch.	Must
WFR.03.10	Multiple assets	The system supports the use of more than one digital asset with one single metadata record.	Must
WFR.03.11	Defining media types	The metadata and media types prescribed by the target are defined automatically on record level or per batch.	Must
WFR.03.12	Metadata field on IPR digital object	The system adds missing or corrected information on the IPR of the digital object based on input of the user manually or in batch.	Must
WFR.03.13	Metadata field on IPR metadata	The system adds missing/corrected information on the IPR of the metadata based on input of the user manually or in batch.	Must
WFR.03.14	Metadata field on IPR preview	The system adds missing or corrected information on the IPR of the preview (thumbnail) based on input of the user manually or in batch.	Must
WFR.03.15	Mark mandatory fields	The system indicates which fields are mandatory for a chosen mapping or output data.	Must
WFR.03.16	Choosing a default mapping	The system supports choosing a default mapping based on user input or system configuration.	Must
WFR.03.17	Automatic data suggestion	The system suggests necessary data enhancements on data set (like apply license, apply source institution) and gives the possibility to approve or decline them.	Should
WFR.03.18	Target format selection	The content provider points out what source format the data is in and chooses a target format.	Should
WFR.03.19	Semantic data enrichment	The system can be used to make data more explicitly semantic by linking or converting data to controlled vocabularies and thesaurus concepts.	Should
WFR.03.20	Conditional mapping	The system supports conditional mappings. The decision about which target field for some content may depend on the value in certain fields.	Must
WFR.03.21	Nested or grouped mapping	The system can perform mappings that consider the structure of nested or grouped fields.	Must
WFR.03.22	Intermediate format mapping	The system can support sequential application of various mappings, e.g. native data model into LIDO into EDM.	Should

WFR.03.23	Support for conditional truncation	The system can truncate the content of certain fields based on predefined conditions (cases).	Could
WFR.03.24	Apply PID	The system must check local identifiers in source data and enhance them automatically for global use based on configurations of the relevant CP.	Must
WFR.03.25	Conditional field conversion	The system can automatically convert certain data values based on predefined conditions. E.g. when [type] = “production place” THEN [eventType] = “Production”).	Could

Validate transformation and receive feedback – Requirements Overview

No.	Requirement	Explanation	Priority
WFR.04.01	Validation	The system validates mapping results against chosen target schema, e.g. EDM.	Must
WFR.04.02	Feedback on validation	The system reports on the irregularities of the mapping results (e.g. missing fields, missing thumbnails).	Must
WFR.04.03	Edit invalidated fields	I think this actually means that if corrections are made then it should be possible to only reprocess these items rather than the whole set.	Must
WFR.04.04	Automatic license validation	License information is validated automatically.	Must
WFR.04.05	Test ingestion	The system is able to do a test ingestion for metadata prepared for ingestion by Europeana.	Should
WFR.04.06	Align validation	The system ensures that successful validation warrants validation by Europeana at ingestion as well.	Must

Minutes Group 3: Supply, Acceptance

Participants: Neil Smith (leader), Carolien Fokke (notes), Michael Selway, Sam Alloing, Sam Coppens, Sašo Zagoranski, Spiros Trivizas, Stefan Rohde-Enslin, Vincent Leconte, Zoltan Mazula, Antoine Isaac, Arthur Hanselman.

Neil Smith (NS): These functional areas are closest to the actual interface with Europeana. Which of these areas can we develop as part of iteration 1? Also, do we develop them together or each vendor separately, or with a subset of vendors?

Supply

There are 2 'MUST' functionalities under Supply, WFR.05.01 and WFR.05.02.

WFR.05.01

Michael Selway (MSL) said that one model is to make your system do it. Then you go per record and you say 'yes, put on Europeana'. He said he envisages a zero-click requirement.

NS replied that people want to be able to have it both automatic and part of a selection process. So in fact, there are two requirements in this single one. They want it to be partly automatic, but they do want to be able to select as well.

NS said that OAI-PMH is already defined. He would like optional parts of support for selection, sets, etc. In the future it will be useful. MSL then replied that he would like an OAI-PMH that is actually push (and not pull). He suggested having a little side strand to make a variant of OAI-PMH which pushes. NS said that there are existing protocols such as SWORD that can do a push. However, we will have to ask Europeana.

Antoine Isaac (AI) from Europeana then said that although Europeana has thought about it, they have not yet taken any decisions about it. NS asked whether it would be possible to have a subset of CPs for data push. Like an ongoing strand. AI said that there is the possibility to discuss a pilot for data push back to Europeana.

NS asked whether enriched content should be pushed back or pulled? For example, in the CultureGrid: push.

NS then said that the only thing we can do at the moment is to use OAI-PMH in iteration 1. MSL asked whether FTP can be used? NS said yes, but Europeana pulls. Marco Streefkerk (MS) suggested the use of SOW, which NS said is a search protocol. MS said it was used as push in the cultural domain in the Netherlands.

NS said that harvesting is difficult and annoying especially because of the conformance checks. There is a system from Capetown that may be of help here, they should be contacted. If we are saying we are going to use OAI-PMH, we have to create our own test harness.

MS said that if Europeana accepts records directly, this may be a lot of work for them. NS said that if all OAI-PMH systems would be better-behaved it would be ok. He then stated that to clarify things, the requirement should be split into a pull requirement and a push requirement. MSL said that the requirement is to be able to use one protocol.

WFR.05.02

MSL said this requirement focused on a technical detail and is protocol-dependent and questioned whether this was really a requirement.

MS explained that this requirement comes from the CPs and the experience they have with the variety of steps they have to go through once there is an error and they have to re-do everything. NS encouraged everyone to think about how this requirement could be turned into a technical specification. AI wanted to know what was considered as a 'fail' in case of a pull. NS explained that this could be a timeout-fail, or for example when there is something wrong with a record. MSL then summarised the requirement as not having to go back to step 1 of a process. He then said this required the system to get status information back from Europeana.

Sam Coppens (SC) from iMinds wanted to know: What is the information Europeana can give back? Can they not say which records failed? Spiros Trivizas (ST) from PostScriptum wanted to know how many times the system would try again before giving an error with some information about the reason for error. Will it try forever, or will it try 3 times and then stop? NS said that sometimes it is good to be able to only change the 7 records that failed. SC then said that in that case it is important for Europeana to know which records have been changed after failure because otherwise they will attempt to harvest the whole set again.

MSL asked whether this requirement is still worth having in the requirements list, even though it is problematic. MS and NS said that yes, it was still a requirement. NS suggested the possibility of log feedback. What feedback does the CP get back for a failed harvest? AI suggested that this could just be a ring. MSL suggested implementing something to tick once Europeana calls that it went wrong. NS said it would be nice if Europeana could receive a request query like in the Culture Grid. AI said that he does not see anything happening that is more than 'this many records have failed'. He thinks the requirement is risky and quite impossible.

Arthur Hanselman (AH), the representative for Gallery Systems, then said that Europeana should be able to give useful information about this. He asked whether Europeana was leading, or whether the consortium was. What if Europeana cannot supply the things that are needed for this 'MUST' requirement? NS said that it was important to discuss things like this with Europeana, but also to go back to the people who wrote this requirement down to find out exactly what it is they want. MS suggested that maybe it would be ok if it worked with an aggregator.

WFR.05.03

OK.

WFR.05.04

MSL said he was a bit nervous that we are building a very general purpose system. That we should focus on something that provides to Europeana. AI suggested that it could be a higher level requirement.

Supply (push / pull) – Requirements Overview

No.	Requirement	Explanation	Priority
WFR.05.01	Automatic supply	The system supplies prepared and validated data to Europeana by push or pull.	Must
WFR.05.02	Re-supply functionality for failed records	In case of an error the system is able to start the supply process again only for the failed records.	Must
WFR.05.03	Schedule data supply	The system can be scheduled to supply data at a predefined date/time.	Should
WFR.05.04	Tools for third-party collaboration	The system facilitates the supply of data to platforms other than Europeana as well and provides the necessary tools (e.g. licensing filters and query APIs).	Could

Data Acceptance

WFR.06.01

NS asked whether this requirement is already part of MINT. AI said he was not sure whether it is a preview really. SC explained that there are previews of the HTML representations. AH said this was quite Europeana focused, and asked whether it would be updated? AI explained that Europeana has had this before. MSL said that the requirement needs to be specialised and focused on whether you have to save something externally before or after it has been sent.

WFR.06.02

AI said that this will not happen and that it is codified in the agreement that CPs sign. NS asked whether the situation would be different if the OAI-PMH can do a deletion (persistent) and put a delete function in there. MSL wanted to know whether it had to be 'instantly'. SC said this could only be achieved via a push. AI suggested it could also happen by doing a re-harvest (of which the process could be speeded up by sending an email to Europeana).

WFR.06.03

ST wanted to know whether this would be done manually or automatically. NS said it should be both. MSL said there should be an ID on it to tell Europeana that it is the same object. SC said the update does not have to be incremental. AI said that you may want to keep track of every record. It could be done by republishing as new record. NS said that that would be easiest.

WFR.06.04 and WFR.06.05

MSL said this was difficult to tackle. This is an open question about communications back channel. He suggested it is necessary to sit down with a group. Some things simply have to be done by email. MS stressed that the requirement was logical from the CPs perspective. Meeting the requirements may also be done in a different way if the one suggested here is not doable. NS said that it was important to still talk to Europeana about many of these requirements.

Data acceptance – Requirements Overview

No.	Requirement	Explanation	Priority
WFR.06.01	Preview presentation Europeana	The system is able to preview the data representation in Europeana before it's being published.	Must
WFR.06.02	Withdraw records	The system can withdraw earlier delivered records instantly from Europeana by instructions of the involved CP.	Should
WFR.06.03	Update published records	The system can keep the data that are already in Europeana up-to-date.	Must
WFR.06.04	Publication indication	The system gives an indication about the processing steps and scheduling in Europeana.	Should
WFR.06.05	Automatic publication alert	The CP is informed on publication of the data on the target website (Europeana or aggregator).	Must

Wrap-up

Much of the management stuff is for iteration 2. It would be nice to see the reason why many records suddenly fail due to an update.
NS then asked which requirements should be prioritised and described for iteration 1.

Answer all: WFR.05.01, WFR.05.02 (with clarification from CPs) WFR.06.01 (could be considered if the API is there. Must look at MINT and use that, clarification from Europeana needed as well), and WFR.06.03.

ACTION: list of questions for Europeana to be sent to Breandán Knowlton.