Grant Agreement 297292

EUROPEANA INSIDE

Minutes of 3rd Technical Partners meeting

Document number	*D1.7*
Dissemination level	*Public*
Delivery date	*January 2014*
Status	*Final*
Author(s)	*Anne Overbeck (SPK)*

Revision History

Revision	Date	Author	Organisation	Description
V 0.1	23.1.2014	Anne Overbeck Monika Hagedorn-Saupe	SPK	Draft Version 1
V 0.2	29.1.2014	Neil Smith	K-INT	Draft Version 1
V 0.3	31.1.2014	Nathalie Poot	KMKG	Draft Version 1
V 0.2	6.2.2014	Phill Purdy	CT	Draft Version 2
V 1.0	7.2.2014	Anne Overbeck	SPK	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This deliverable contains original unpublished work except where clearly indicated otherwise.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 3rd Technical Partners meeting
Place and Date Maribor, 11-12th of December 2014 **Venue** Museum of National Liberation Maribor	

This is the 3rd Technical Partner Meeting of the Europeana Inside project. Its aims are:

- Review the project's technical progress to date;
- Identify areas where existing ECK components can be improved;
- Plan activities for work package 5, including developing new functionality in iteration 3;
- Plan testing of iteration 4;
- Discuss options for maintaining ECK components post-project.

Participants

Name	Organisation	Country	Email
Sam Alloing	Catholic University of Leuven	BE	sam.alloing@libis.kuleuven.be
Ágoston Berger	Monguz	HU	aberger@monguz.hu
Eric De Cacqueray	Mobydoc	FR	eric.de-Cacqueray@mobydoc.fr
Laszlo Karman	Monguz	HU	lkarman@monguz.hu
Breandan Knowlton (Dec 12th from 9-9:45)	Europeana	NL	breandan.knowlton@kb.nl
Jörg Kruschinski	Zetcom	DE	Joerg.kruschinski@zetcom.com
Naeem Muhammad	Catholic University of Leuven	BE	Naeem.Muhammad@libis.kuleuven.be
Anne Overbeck	Stiftung Preussischer Kulturbesitz	DE	a.overbeck@smb.spk-berlin.de
Nathalie Poot	Royal Museums of Art and History	BE	n.poot@kmkg-mrah.be
Phill Purdy	Collections Trust	UK	phill@collectionstrust.org.uk
Neil Smith	Knowledge Integration Ltd	UK	neil.smith@k-int.com
Marco Streefkerk	Digitaal Erfgoed Nederland	NL	marco.streefkerk@den.nl
Michael Selway	System Simulation Ltd	UK	mas@ssl.co.uk
Martijn Van der Bruggen	Adlib	NL	m.vdbruggen@adlibsoft.com
Chas Woodfield	Knowledge Integration Ltd	GB	chas.woodfield@k-int.com
Sašo Zagoranski	Semantika	SI	saso.zagoranski@semantika.si
Nina Zagoranski	Semantika	SI	nina.zagoranski@semantika.si
Tadej Kokovnik	Semantika	SI	Tadej.kokovnik@semantika.si

Not in attendance:
- KE
- Skinsoft
- Postscriptum
- iMinds

Agenda

11th of December

Agenda item	Minute
Welcome and Tour of the Table	Phill Purdy (PP) from Collections Trust introduced himself as the new project manager, welcomed everyone and everyone in attendance introduced themselves shortly.
14:00-17:30 Practical Session	Background: Prior to the meeting, every partner had been asked to identify a small set of around 10 test records and use ECK functionality to convert these to LIDO using PID generation, preview services, etc. according to their workflow and note any issues they encounter. Each partner was asked to bring the records on a USB memory stick or email them in advance to chas.woodfield@k-int.com. During the session, a set of LIDO records from each technical partner were uploaded into the dark aggregator (DA) to 1) check the LIDO form and 2) to revise the transformation to EDM. Any issues that occurred during the process were discussed with the plenary.
14:00-16:30 Practical Session, Part 1	Neil Smith (NS) encouraged TPs to submit as much data as possible to enable the team to identify problems early. NS told that validation was only done against LIDO not EDM. NS reported the three ways of submitting data to the dark aggregator: SWORD, web form and OAI-PMH. Chas Woodfield (CW) introduced the management portal and showed how to log in by using the company name. He stated that both the Dark Aggregator (DA) and Culture Grid (CG) were based on the same, existing, open source product developed by K-INT. CW explained that, in the management interface, the LIDO_raw and EDM versions of a record were relevant for this project. The web form interface accepts either a single LIDO record, multiple LIDO records in a single file (using lidoWrap) or a zip file containing multiple records. CW encouraged TPs to pass bugs on to K-Int. Sum up of live testing: **ADLIB** 137 records were uploaded in the DA, all were invalid. Validation error (Semantika): the element *titleWrap* in *namespace* 'http://www.lido-schema.org' has incomplete content. List of possible elements expected: 'titleSet' in namespace 'http://www.lido-schema.org'. There is only *titleWrap* in the LIDO, which needs to be: ObjectIdentificationWrap<TitleWrap<TitleSet<AppelationValue. All records are missing a title, which is a mandatory field in LIDO. **Action Point:** Adlib to correct the LIDO xml. Sam Alloing (SA) from LIBIS asked about how to deal with multiple images. Chas Woodfield (CW) from K-Int answered that the DA calculates the larges, if there are no measurements, it takes the first one. SA stated that *edm:object* is the important field for this, not *edm:shownby*. **K-INT** 4 records were uploaded in the DA, they were all valid.

The following LIDO fields did not show in EDM:
- Lido:ObjectNote(Lido:relatedWorksWrap<lido:relatedWorkSet<lido:relatedWork<lido:object<lido:objectNote)
- Lido:nameActorSet<lido:appellationvalue (lido:eventWrap<lido:eventSet<lido:event<lido:eventActor<lido:actorinRole<lido:actor<lido:nameActorSet<lido:appellation value)

NS stated that could be multiple fields going to description, but that their validation did not take this into account.

NS stated that the current LIDO to EDM transformation was based on XSLT supplied by the Linked Heritage project. The plan is to review and refine this in future iterations.

CW addressed issues regarding the EDM field *dc:identifier*. He stated that Europeana does not use it as unique identifier. Instead, when you get that record back, it does have a new identifier, given to the record by Europeana.

Nathalie Poot (NP) stated that the *lidoRecID* goes into *rdf:about* in EDM.

MOBYDOC
2 records were uploaded in the DA, they were both invalid.

Validation error (Semantika): The element *administrativeMetadata* in namespace 'http://www.lido-schema.org' has invalid child element *rightsWorkSet* in namespace 'http://www.lido-schema.org'. List of possible elements expected: *rightsWorkWrap, recordWrap* in namespace 'http://www.lido-schema.org'.

Lido:rightsWorkWrap is missing.
- <u>Now</u>: Lido:administrative metadata <lido:rightsWorkSet<lido:rightsHolder<…
- <u>It should be</u>:Lido:administrative metadata *<lido:rightsWorkWrap*<lido:rightsWorkSet<lido:rightsHolder<…

For the other record there was no validation error mentioned. It was only stated 'validation failed'.

Action Point: Mobydoc to correct the LIDO xml.
Action Point: Monguz and Semantika to take a look at the validation modules.

MONGUZ
10 records were uploaded in the DA, they were all valid.
conceptID and *lido:term* were not picked up in EDM.
lido:rightsWorkWrap was missing

SEMANTIKA
11 records were uploaded in the DA, 6 were invalid.
Different validation results:
- Validation Monguz: failed (not clear why validation failed. No explanation)
- Validation Semantika: successful

Action Point: Ágoston Berger (Monguz) and Sašo Zagoranski (Semantika) to take a look at the validation modules.

	SSL
	Four records were uploaded in the DA, they were all valid.
	Lido:descriptiveNoteValue and *dc:description* (EDM) is not taken from a controlled vocabulary.
	
	ZETCOM
	20 records were uploaded in the DA, they were all valid.
	The mulitple *Dc:identifiers* are taken from:
	   • Lido:lidoRecID (2 lido:lidoRecIDs)
	   • Lido:lidoPublishedID)
	   • Lido:recordID
	   • Lido:WorkID
	
	NS stated that there should only be one *lidoRecId* to avoid problem of having the wrong one ending up in Europeana.
	
	Jörg Kruschinski (JK) from Zetcom said that to identify one lidoRecId is very important and that the TPs should agree on one type.
	
	CW said that that Europeana takes the unique identifier from the field *edm:ProvidedCHO* or *rdf:about* and not from *dc:identifier*, since it is possible to have multiple identifiers.
	
	Action Point: CW pointed out that all TPs can continue testing in the test environment of the DA. There is a login for each TP with which they can check their own records.

16:30-17:30 Practical Session, Part 2	**Lead Question: Should the two validation modules return identical results?** The validation modules validate the LIDO xml against the schema and the form. The validation module should be part of the CMS of the CPs. The two validation modules (by Monguz and Semantika) present at times different results (for example records are not transformed to EDM, because there are errors according to Monguz, but when clicking on the record itself to validate, the validation is deemed successful by the module from Semantika. When there are several errors in one LIDO records, only one error is shown. Sam Alloing from LIBIS said that Europeana showed that they were validating with schematron. Agoston Berger (AB) from Monguz said that he would check this. **Action Point:** Both modules should present the same result, so that when one is down, the other can be used. Monguz and Semantika will look into it. **Lead Question: Should we be validating EDM as well as, or instead of, LIDO?** NS stated that at the moment there is only preview based on LIDO. He posed the question whether there should be preview based on EDM. Michael Selway (MSe) stated that at the beginning of the project the choice was made to validate and preview LIDO, not EDM. By adding an extra preview and validation, this will complicate the workflow process to Europeana, while the idea of

	Europeana Inside is to simplify the process. SA from LIBIS argued that a preview based on EDM is necessary, since Europeana takes EDM for harvesting. On the other hand, NS and CW argued that if the LIDO-EDM transformation is correct, the preview should present the same result. CW added that the project's aim is to make sure that it is right on Europeana. CPs and TPs may want to pass it on elsewhere. In that case one might want to validate not to EDM, but to LIDO. JK asked what the EDM feedback meant to CPs. CPs want to see end-result in their own CMS. All CPs can export LIDO, but not all EDM (transformation LIDO-EDM is part of the DA). When there is only a preview based on EDM, most of the CPs will not be able to see it. NS stated that Antoine Isaac from Europeana has a strong interest in EDM validation. MS said EDM validation should come from Europeana, otherwise, TPs would have to depend on Europeana to communicate changes. **Lead Question: Should the preview module work on the EDM rather than the LIDO?** MS raised the question whether the preview should work based on EDM rather than LIDO. NS said that his should not make a difference. AB stated that he started with a preview based on EDM. SA stated that it needs to be developed for EDM, because LIDO is only an intermediate. JK said that he only wants to deliver LIDO to the preview service, but that the preview service should transform it "under cover" to EDM and check it. NS said that CPs do not care about the format as long as it looks right on Europeana. **Action Point:** It was concluded that Europeana should be asked about an EDM validator and the decision about how to deal with this within the Europeana Inside Project was postponed until then. **Action Point:** Phill will plan a meeting in January 2014 with Europeana to discuss the validation to EDM. Europeana themselves are working on EDM validation.

Agenda

12th of December

Agenda item	Minute
09:00-09:15 Introduction (Phill Purdy)	PP gave an update on the status on the project, upcoming test phases, deliverables, meetings and networking event. PP stated that in Period 1 all deliverables and milestones were achieved and a successful review by the commission had taken place.

	PP continued in saying that in Period 2 all deliverables and milestones were achieved. Review for this period were in progress
	PP listed Period 2 successes to be: successful conducting of Iteration 2, good progress on CMS integration, good growth in Associate Partners, close collaboration between Content Partners & Technical Partners during testing period, the submission of 85,000 records from 3 collections to Europeana, and the completion of Period 2 Reporting.
	PP reminded the TP to keep Period 1 Review recommendations in mind. These were: a) simplifying the metadata provisioning, enrichment and re-integration for content providers and b) demonstrating this progress by substantially increasing the amount of content made available to Europeana.
	Finally PP listed the tasks that lie ahead concerning the final iterations of ECK, more content ingestion to aggregators and Europeana, the 3rd Network Event in Athens in April, the 4th Technical Partner meeting in Toulouse in July and the Launch Event in London in September 2014.
09:15-09:30 Review of iteration 2 testing (Nathalie Poot)	NP reported that in preparation of testing iteration 2, additional meetings were held between technical partners and content providers. The meetings took place in Budapest, Athens and Leuven. This gave technical partners the opportunity to present their test plan for iteration 2 and content providers the possibility to ask questions on how the testing needed to be performed.
	NP went through the forms used during the testing process of iteration 2. She stated that all CPs completed the Acceptance Test Form i2. The answers can be reviewed in *D4.1(v2) Control Export Evaluation Report.* She stated that there were slight differences in the way CPs interpreted 'not accepted' and 'not tested'. Some CPs indicated that a FR is 'not accepted' while they were not able to test it, due to a shortage of time.
	NP reported that the majority of FRs that were 'not tested' belonged to Supply and Data Acceptance. She stated that most of those FRs are incomplete, because they depend on the implementation of these functionalities by Europeana.
	NP said that for some FRs that were accepted, CPs commented that they accepted the FR because it is present and works, but that it is too complicated to be performed by the CPs themselves.
	Action Point: For the upcoming iterations, improvements are needed on usability.
	NP told that the goal of the Content Providers Survey i2 was to evaluate the test process and not the functionalities of the ECK. She stated that content partners provided feedback.
	NP stated that the accessibility of the test instance improved for several content partners improved in comparison to testing iteration. All content partners were now able to access their test instances easily.
	NP reported that all partners indicated that they received good technical assistance. However not all partners received documentation on the test instance and did not fully understand how the functionalities needed to be tested. Since time for testing was limited, there was not always enough time left to ask for clarifications. She also stated that time restraints made the complete testing of iteration 2 difficult.
	NP reported that several content partners made data available to the (dark) aggregator by OAI-PMH or uploaded their records by web form.

	NP concluded in saying that the overall evaluation of ECK iteration 2 is good. The majority of the content partners gave a 'good' to 'very good' evaluation. A more negative evaluation refers primarily to the test process (not enough time for testing and no documentation) and the usability (not all accepted FRs can be easily performed). NP stated that the iteration showed that there was not enough time allocated to testing. She said that the fact that the new functionalities needed to be installed, tested and evaluated in one month influenced the test results. NP informed that the commercial vendors of collection management systems without a direct CP within the consortium had difficulty in finding a testing partner. NP pointed out that this made it difficult to evaluate what these partners have developed.
09:30-09:45 Data Push (Neil Smith)	NS gave a status update on WP5. He stated that the goal of the project is to have the ECK implemented in the CMSs of the CPs in June 2014. NS stated that changes were made to the DOW due to differences in the development plan and that it was permissible to move PM from WP3 to WP5. He reported that a fair amount of resources were left for two things: fine tuning and two new areas of functionality. These were data push and content re-ingestion. NS told that the Integration Status Report in August will need to state that the ECK has been integrated in each participating CMS, repository platform or service. **Action point:** NS said that at the 3rd Networking Event in Athens it should be possible to show to the CPs how the modules will be integrated. NS said that the aim of iteration 3 would be to continue the development of components of iteration 1 and 2, and to test data push and content re-ingestion.NS said that the aim of iteration 4 is to fine-tune the existing functionalities but would include no new functionality NS stated there are two options for the implementation of data push: TPs can either implement OAI-PMH or data push (SWORD). Concerning Data Push, NS summed up the Current Position of Europeana. He said that supply to Europeana was possible via OAI-PMH Repository (for harvesting) or by exporting a data file to an FTP site. NS stated that from a vendor's view export to FTP site is a poor solution because it is impossible to know when it is being picked up and no feedback is given. He stated that maintaining an OAI-PMH Repository is difficult, because this method needs firewall configuration, the harvesting can place a load on the system as there is no incremental harvesting by Europeana and also here it is impossible to know when it is being picked up and no feedback is given. NS said that even after repeated requests, Europeana is silent on when this question will be picked up. NS summed up the advantages to be that the data provider controls timing, the data provider knows when data has reached the target (no firewall configuration required) and no need to maintain public facing repository NS stated that SWORD had been implemented in the DA and can be tested in the DA. NS stated that CT and K-Int are lobbying for museum sector aggregator to accept data push via SWORD. **Action point:** CW stated that he will provide documentation for testing.

	NS strongly encouraged TPs and CPs to make their voices heard through the list, if they liked data push. He stated that dissent was coming only from big national institutions that submit data to Europeana directly

Breandan Knwolton (BK) from Europeana stated that data push is not the same as final publication. He stated that everything sounded feasible. But that Europeana had to figure out how to do it. He stated that Europeana was aiming at getting the process of updating data down from 4 weeks to 2 weeks by the middle of the year, but that this was not working yet. BK excused himself from the discussion because he had to catch a train.

Action point: CW suggested making SWORD the focus of the TP call at the end of January.

Action point: PP is planning a meeting with Europeana in January 2014 to discuss the issues surrounding data-push and content re-ingestion. |
| 09:45-10:30 Content Re-ingestion (Chas Woodfield) | CW gave an update on the content re-ingestion. He stated that the enriched content refers to: subjects, location and actors (creation).The lidoRecID will be taken to match the records.

CW presented a seven step process on how to perform data enrichment, but pointed out that there is no need for the suppliers to go through these steps unless they have a desire to do so, as the aggregators will do the job for them and make the data available to them in a sensible form. A proposal on the specifics of this steps has been send to the Europeana Tech list.

CW pointed out that the option of the enriched records being made available in an OAI server was mentioned, but subsequent queries about this to Europeana have neither been denied nor confirmed and there was no mention of a timeframe.

MSe pointed out that if content reingestion is not possible due to development issues on the side of Europeana, this should be communicated to the PO. He stressed that this would need to be communicated very clearly.

SA stated that Europeana gets from these discussions what they should do, but does not take steps to act.

MS said that it should be possible to show clients /CPs what content ingestion can do.

MSe stated that development in this area is very interesting from the viewpoint of open data.

Martijn Van der Bruggen (MB) and JK said that it is important that the CP stay in control and have a choice whether or not to accept enriched data.

CW said it is up to the CMS to decide what to do with the enrichment.

MB raised the question of how to match the new record ID provided by Europeana with the old record ID given by the CMS after enrichment by Europeana.

CW stated that right now in order to get an item back from the aggregator one needs to know the new ID. CW stated that K-Int can find that out via lidoRecId and that this works on an individual record basis.

MB stated that it does not make sense to submit something, if a CP loses hold on their own records. |

	MSe repeated the importance of including the PO in the decisions making process and informing them on what is going wrong.
	Action point: The lack of cooperation from Europeana raised questions with all TPs to the feasibility of the requirements. All the FRs related to content re-ingestion in D2.4 are 'should' or 'could', none of them are 'must'. A solution might be to develop methods of content enrichment but not actually implement them. This change should be communicated to the EC beforehand since content re-ingestion is part of the DoW and the project will be evaluated on it.
	Action point: PP is planning a meeting with Europeana in January 2014 to discuss this.
11:00-11:30 Iteration 4 – testing and final deliverables (Neil Smith & Nathalie Poot)	NP and NS presented the iterative development plan for iteration 2. NP proposed a change in timing. NP suggested releasing the iteration 4 production version one month early, in M27/June 2014. To allow more time for testing and feedback in M28/July 2014 and reporting in M29/August 2014. The TPs agreed to the changes. NP reported that while the evaluation forms for iteration 1 and 2 focussed on the presence of functional requirements (CPs indicated that FRs were 'accepted', 'not accepted' or 'not tested') iteration 4 will focus more on usability. The evaluation forms will be adjusted accordingly. NP pointed out that by June 2014 several CPs will have published all their content on Europeana. NP said that test records will be foreseen to participate in testing of iteration 4.
11:30-12:30 Post project – forward plan, availability of modules, updating (Phill Purdy & Marco Streefkerk)	PP presented that in order to secure the viability of the Europeana Inside BPN and its outputs, a Forward Plan will be developed to address the future development potential of the Europeana Connection Kit, the maintenance of the Best Practice Network, and the promotion and advocacy of the Europeana Inside outputs to new partners. PP summed up what had been achieved so far and outlined the next steps. PP said that getting input from TPs was crucial for the development of the Forward Plan to figure out questions about what key business drivers / barriers to take up and to identify any key features from user-requirements that could be added in next iterations to support take up e.g. showing supply to other end-point(s). Marko Streefkerk (MS) from DEN presented the Business Model Innovation Cultural Heritage Roadmap. MS pointed out that the ECK modules are developed by different technical partners and are being maintained on their servers. After the project it will be necessary for the modules to stay available. He proposed the Possibility to assemble all functionalities and to assign them to Europeana to store them centrally. MS concluded in saying that the details of this will be discussed in the upcoming months. In the ensuing discussion PP pointed out that the project was aiming to be a long term solution also aiming at recipients other than Europeana. MSe pointed out that museums need to want to get their data into Europeana, but expressed doubts about that was something, Europeana Inside could achieve. He said that until museums want this themselves, being told to do it by national

institutions is good way to get things done.

MB pointed out that CPs will continue to be reluctant to contribute to Europeana as long as quality of the data in Europeana does not improve.

MS said that there is not a lot of demand to give data to Europeana, but that there is demand to open up content and CPs want that.

JK said that in his experience CPs start out with their own website, but want to go to Europeana. He said that it is not about building a sexy platform but about being part of the community. Therefore, to him, the long term view and a reliable infrastructure is very important. He said that it is the TPs own responsibility to update their own components of the ECK. He reported that Postscriptum proposed a platform for all services.

SZ expressed the hope that services will be centralized in Europeana.

AB stressed this by saying this would be the best way to keep the systems running.

MS expressed the possibility that Europeana might be able to support this because they have public funding and that he hopes for constant funding for this hosting service.

MB said that he supports the solution to put the services all in one central place as close to Europeana as possible and that updates should be part of the Europeana onward development.

After the presentation and the discussion PP announced the plan to start a working group on sustainability and asked for TPs to join the group. Monguz stated that they might be interested.

Action point: PP announced that he will get in touch with them individually about the creation of a working group on sustainability.

Toulouse, France, was decided to be the place of the next TP meeting.

PP said that the main aim of the next months would be to refine the LIDO to EDM process and reminded the TPs that monthly and quarterly reports were required.
PP concluded the meeting.