Europeana – Core Service Platform

DELIVERABLE

D1.3: DESCRIPTION OF PLANNED WORK FOR THE AGGREGATORS ON IMPROVING THE DATA

Revision	1.0
Date of submission	31 July 2015
Author(s)	Henning Scholz (EF), Julia Welter (DIF), Maria Teresa Natale (MICHAEL), Marzia Piccininno (MICHAEL), Corinne Szteinsznaider (MICHAEL), Walter Berendsohn (FUB), Gisela Baumann (FUB), Kate Fernie (2Culture), Dimitris Gavrilis (AthenaRC), Marco Rendina (eFashion), Johan Oomen (NISV), Gariella Ivacs (IALHI), Anastasia Gasia (EF), Valentine Charles (EF)
Dissemination Level	Public

Funded by
the Connecting Europe Programme
of the European Union

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author (Organisation)	Description
0.1	14/07/2015	Henning Scholz (EF)	Template and initial draft
0.2	27/07/2015	Henning Scholz (EF), Julia Welter (DIF), Maria Teresa Natale (MICHAEL), Marzia Piccininno (MICHAEL), Corinne Szteinsznaider (MICHAEL), Walter Berendsohn (FUB), Gisela Baumann (FUB), Kate Fernie (2Culture), Dimitris Gavrilis (AthenaRC), Marco Rendina (eFashion), Johan Oomen (NISV), Gariella Ivacs (IALHI), Anastasia Gasia (EF), Valentine Charles (EF)	Final draft
0.3	30/07/2015	Gordon McKenna (CT), Marie-Claire Dangerfield (EF)	Review
1.0	03/08/2015	Henning Scholz (EF)	Reviewers comments integrated, layout edits

Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise.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

Table of Contents

Introduction

The variety of content that is accessible via the Europeana Portal is huge. Books, paintings, photographs, newsreels, sound recordings from more than 3,300 museums, libraries, archives and audiovisual archives from across Europe have been made accessible over the last few years, resulting in a corpus of more than 44.7 million metadata records (as of 13 July 2015). But not only the content itself is very diverse, the quality of the data a user can find in Europeana varies a lot across the Europeana database. When we say 'data' we mean both the metadata describing the cultural heritage objects and the digital objects linked from within the metadata. For more details about metadata quality in Europeana it is referred to the Report and Recommendations from the Task Force on Metadata Quality[1].

Improving data quality is the first priority for Europeana under the Europeana Strategy 2015-2020. It is clear from the work of the Metadata Quality Task Force how important the contribution of aggregators towards better data quality is. We see that aggregators that invest in more metadata checks achieve a higher quality compared to aggregators investing less in metadata checks. A much closer collaboration between Europeana Office and aggregators but also between different aggregators creates a better understanding of the needs for better data and how to achieve this. The set-up of the Europeana DSI is supporting this closer collaboration. Eight domain and thematic aggregators are full partners of the Europeana DSI to work more closely together on various aspects of aggregation, including data quality improvements. The purpose of this deliverable is to make plans for data improvements transparent between all aggregating partners of the DSI to enable a better alignment of activities and to address specific challenges together, using the experiences of other aggregators. In setting up this plan it also helps to advice aggregators on priority targets and potential risks of choosing a certain approach.

The following documents are meant to be a guide for good targets for quality improvements:
- Europeana Publishing Framework (to be published later in 2015)
- Europeana Publishing Guide[2]
- Partner and Data Development Plan[3]
- Report and Recommendations from the Metadata Quality Task Force[4]

A few examples are given to illustrate potential targets for quality improvements, that are elaborated in more detail in the above documents. If aggregators want people to find, view and use content, however users want to do that, content has to be freely re-usable (PDM, CC0, CC BY, CC BY-SA), directly linked from the metadata and the technical quality needs to be high (images need to be 1,200 pixels wide, which is equivalent to about 1 megapixel). Enrichment is an important element that improves data quality, e.g. through the use of open vocabularies such as Iconclass and Getty Art and Architecture Thesaurus. Thumbnails (or other previews) should be made available for all digital cultural heritage objects. Metadata should have relevant, human-understandable fields, e.g. titles should provide meaningful access for the users of Europeana. Depending on the quality of existing data in Europeana, the aggregators focus their efforts on specific elements to improve their collections, which is specified in this document.

Although quality targets should be measurable as much as possible, this document is not to elaborate on the new data to be ingested. For new data ingestions it is important to follow the above guides and recommendations, but the amount of data will only be presented in September 2015 (Europeana DSI MS6 - Data ingestion plan). The focus of the quality improvements listed and described below are the data submitted by the aggregators in earlier years.

---

[1] http://pro.europeana.eu/publication/metadata-quality-task-force-report
[2] http://pro.europeana.eu/publication/publication-policy
[3] http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Version3/Milestones/Ev3%20MS2%20MS4%20MS5%20Partner%20and%20Data%20Development.pdf
[4] http://pro.europeana.eu/publication/metadata-quality-task-force-report

Planned work for the aggregators

1-European Film Gateway (EFG)

Person Months in Task	6

Description of work
EFG will clean and enrich metadata coming from the film archive domain to EFG and Europeana. The focus of enrichment work will be on improving discoverability of digital objects by enriching keywords and content descriptions. To support the user experience on EFG and Europeana, EFG archives will be encouraged to provide direct links to digital objects and thumbnails, where this is not already the case, and to improve the rights labelling of individual objects so that their rights status becomes more transparent to users. Together with the EFG partners from the film archives, Deutsches Filminstitut (DIF) will work on these tasks. In addition, EFG seeks to connect new archives as well as new collections from existing EFG archives.

Subcontracting	
Task (Europeana DSI)	1.5.2 Improve the Europeana DSI data and metadata quality 1.7.2 Develop, establish and maintain technical infrastructure
Amount (EUR)	48.000 EUR
Description	

Description
Tasks to be subcontracted: ● Hosting and maintenance of EFG database; ● Participation in discussion and planning of issues concerning T1.1 "Data and aggregation infrastructure innovation"; ● Ingestion work for new data providers and enriched data sets from existing providers; ● Improvement of tools (e.g. Metadata Editor Tool).

Task	T01.1 Evaluation of possible contributions and data improvements through EFG archives
Description	Association des Cinémathèques Européenes (ACE) and DIF carried out an EFG information meeting to inform all ACE member archives connected or not yet connected to EFG about the activities under the Europeana DSI. The meeting was used to bring attention to data quality issues raised by Europeana under the DSI. As a follow-up of this meeting held in Bologna on the 3rd of July 2015, ACE circulated a survey among all ACE member archives to find out about what they offer to work on under the Europeana DSI, where ACE has reserved a dedicated subcontracting budget for to pay ACE member archives and others for their activities. Deadline for this survey is 14 August 2015. The input provided by the archives will serve as the basis for further negotiations with the archives about concrete tasks to be carried out. At the end of the negotiations, there will be a concrete activities plan for all ACE archives that have indicated their interest to contribute to the Europeana DSI. All further tasks described for EFG in this deliverable depend on the outcome of

	the individual negotiations with the archives. Once the negotiations are finished, it is also possible to provide concrete numbers of objects to be enriched or newly provided.
Start Month	03[5]
End Month	11
Milestone(s)	M01.1 EFG Information meeting held (Month 03) M01.2 Survey circulated (Month 03) M01 3 Surveys evaluated and tasks identified (Month 05) M01.4 Data enrichment and contribution activities by archives to EFG completed (Month 11)
Deliverable(s)	D01.1 Activity plan for EFG archives (Month 07) D01.2 Delivery of improved metadata and new collections to EFG (Month 11)

Task	T01.2 Improvement of rights labeling
Description	Metadata in EFG is completely rights-labelled but not in all cases correctly. This is especially true for contributors who delivered larger amounts of content without rights labels within the original EFG project (2008-2011). In these cases the status was set to the default value "rights protected" for all their content. Seeing that the majority of film-related is still in copyright today, the number of items incorrectly labelled can be considered comparably small. Nevertheless, EFG will aim to specify rights labels more clearly, especially with regard to the public domain. This activity is dependent on the support of the archives contributing the content in question, as re-labelling of rights cannot be done without the individual content provider.
Start Month	04
End Month	11
Milestone(s)	M01.5 Identification of all objects that need relabeling and of archives that have to provide the respective information (Month 04) M01.6 Start of rights labelling (Month 06) M01.7 100% of identified incorrect rights labeling adjusted for correct display on Europeana (Month 13)
Deliverable(s)	D01.3 Rights labels corrected where possible and ingested into EFG (Month 11)

Task	T01.3 Direct links
Description	EFG is already providing the majority of its content with a direct link. Under this task EFG will evaluate the exact numbers of IsSHownBy links provided so far (estimates: c80% of the content coming to Europeana from EFG has an IsShownBy link. c70% of videos aggregated by EFG have a direct link.) EFG will continue to communicate the need for direct links to the archives and engage in increasing the amount of items, especially video, with an IsShownBy link and identify archives that can provide edm:isShownBy in addition to edm:isShownAt.

[5] 00=Apr 2015, 01=May 2015, 02=Jun 2015, 03=Jul 2015, 04=Aug 2015, 05=Sep 2015, 06=Oct 2015, 07=Nov 2015, 08=Dec 2015, 09=Jan 2016, 10=Feb 2016, 11=Mar 2016, 12=Apr 2016, 13=May 2016

	DIF will encourage all contributing archives to deliver edm:isShownBy where archives do not claim strategic reasons for not doing so, or where technical obstacles hinder the delivery of edm:isShownBy. In the latter case, DIF will support archives in finding solutions to overcome these obstacles if possible.
Start Month	04
End Month	11
Milestone(s)	M01.8 Identification of missing edm:isShownBy (Month 04) M01.9 Negotiation with respective archives about possibility to provide edm:isShownBy (Month 05) M01.10 Delivery of edm:isShownBy through involved archives to EFG completed (Month 11)
Deliverable(s)	D01.4 Delivery of edm:isShownBy completed for all EFG archives involved in the ACE subcontract that are able to provide direct links to EFG (Month 11)

Task	T01.4 Providing previews (thumbnails)
Description	ACE and DIF will put a focus on the provision of previews (thumbnails). There are still a number of especially video collections in EFG that do not provide them. In contrast to images and texts where thumbnails can be created automatically by the EFG and Europeana system, this does not work for video files. In the case of videos, the contributing archive has to create thumbnails and deliver them in the form of a link in the exported metadata. DIF will contact the partners in question and encourage them to provide thumbnails if they can.
Start Month	04
End Month	11
Milestone(s)	M01.11 Identification of collections missing thumbnails completed (Month 04) M01.12 Negotiations with archives completed that need to provide thumbnails for videos (Month 06)
Deliverable(s)	D01.5 Delivery of links to thumbnails to EFG completed by archives involved (Month 11)

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
European Film Gateway															
Task															
T01.1	*Evaluation of possible contributions and data improvements through EFG archives*				█	█	█	█	█	█	█	█	█		
Deliverables															
D01.1	*Activity plan for EFG archives*								█						
D01.2	*Delivery of improved metadata and new collections to EFG*												█		
Milestones															
M01.1	*EFG Information meeting held*				█										
M01.2	*Survey circulated*				█										
M01.3	*Surveys evaluated and tasks identified*					█									
M01.4	*Data enrichment and contribution activities by archives to EFG completed*												█		

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
European Film Gateway															
Task															
T01.2	*Improvement of rights labelling*					█	█	█	█	█	█	█	█		
Deliverables															
D01.3	*Rights labels corrected where possible and ingested into EFG*												█		
Milestones															
M01.5	*Identification of objects that need relabeling and of archives that have to provide the respective information*					█									
M01.6	*Start of rights labelling activities*							█							
M01.7	*100% of identified incorrect rights labeling adjusted for correct display on Europeana*														█

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
European Film Gateway															
Task															
T01.3	*Direct links*					■	■	■	■	■	■	■	■		
Deliverables															
D01.4	*Delivery of edm:isShownBy completed for all EFG archives involved in the ACE subcontract that are able to provide direct links to EFG*												■		
Milestones															
M01.8	*Identification of missing edm:isShownBy*					■									
M01.9	*Negotiation with respective archives about possibility to provide edm:isShownBy*						■								
M01.10	*Delivery of edm:isShownBy through involved archives to EFG completed*												■		

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
European Film Gateway															
Task															
T01.4	*Providing previews (thumbnails)*					■	■	■	■	■	■	■	■		
Deliverables															
D01.5	*Delivery of links to thumbnails to EFG completed by archives involved*												■		
Milestones															
M01.11	*Identification of collections missing thumbnails completed*					■									
M01.12	*Negotiations with archives completed that need to provide thumbnails for videos*							■							

2-MUSEU

Person Months in Task	6

Description of work
MUSEU represents three main sources of records for Europeana - the former projects ATHENA, Linked Heritage, and AthenaPlus. MUSEU will guarantee the sustainability of the records of these projects (and their presence in Europeana) over Europeana DSI. Moreover, MUSEU will involve new datasets and data providers within Europeana DSI. The metadata improvement process will be applied to the new records. Former ATHENA, Linked Heritage, and AthenaPlus partners will be involved on a voluntary basis to enrich and refine the data they have already supplied with. MUSEU will address the improvement of the museum data quality, in particular in terms of better visualization and indexing in Europeana. This will be done by promoting the '*Requirements for Visualization and Indexing of Museum Content*' that are under finalization within the AthenaPlus project. Using these requirements for improving the quality of museum data on the provider's side as well as their representation in Europeana will enhance searches for museum items, and hence their accessibility in Europeana. The requirements were created to meet the needs of the museums in the Europeana environment, following the Europeana metadata quality recommendations.[6] They will also empower the user's visual experience through the provision of multiple and better quality images, and also by dramatically improving the quality of metadata with the use of well established vocabularies that will make the retrieval process more accurate. The following activities will take place: ● Definition of the museum sector requirements in terms of expressivity and quality of metadata using LIDO as common framework; ● Implementation of an EDM-LIDO profile in the MINT platform and mapping of the LIDO records to it; ● Integration of the museum requirements in the MINT aggregation platform in order to give content providers the appropriate tool to deliver high quality data; ● Integration of established multilingual thesauri such as the Getty vocabularies AAT and TGN into MINT for empowering the retrieval process; ● Configuration of the Europeana Connection Kit to allow a wider range of museums to submit content to Europeana enriched in their data back into their core systems ● Training and support of content providers for the enrichment of their metadata with the Getty vocabularies AAT, TGN, and ULAN in order to increase the linking of Europeana content to the most relevant LOD sources for the museum sector; ● Training and support of content providers in the mapping process, with a more extensive use of the LIDO elements and also of the normalization functionalities of MINT in order to improve metadata consistency; ● Fostering the application of licenses suitable for the reuse of digital content.

---

[6] See the good practice of the Skoklosters slott - AthenaPlus data provider - in the '*Report and Recommendations from the Task Force on Metadata Quality*'.

Task	T02.1 Museum sector requirements
Description	MICHAEL will revise the museum sector requirements as defined within the AthenaPlus project in order to enhance the accessibility of the museum content in Europeana. The museum sector requirements are structured on the LIDO schema as common framework that allows to show the metadata in their full expressivity. The LIDO to EDM mapping will complete the work and assure the integration in the Europeana workflow.
Start Month	02
End Month	07
Milestone(s)	M02.1 Release of the museum requirements (Month 07)
Deliverable(s)	D02.1 MUSEU data quality museum requirements and LIDO-EDM mapping (Month 07)

Task	T02.2 Metadata improvement
Description	Old and new data providers will supply Europeana with metadata records displaying richer information. This improvement will happen mainly through: ● A more extensive use of the LIDO elements and also of the normalization functionalities of MINT in order to improve metadata consistency and compliance to the Europeana guidelines on metadata quality ● The use of the standard and multilingual vocabularies, in particular the Getty vocabularies such AAT and TGN that empower the retrieval process ● The application of the Europeana IPR licenses suitable for the re-use of digital content ● The application of direct links to the digital objects. These activities will be done using a new and dedicated MINT instance shaped on the basis of the museum sector requirements as described in MS9. MICHAEL will provide through MUSEU-Hub training and helpdesk support for this task as presented in MS8. MICHAEL will ensure that the Europeana Connection Kit (ECK), together with its 'dark aggregator' continues to be supported and potentially be integrated into Europeana's workflow. Quality is ensured for those using the ECK because of its control of metadata (based of SPECTRUM) and in particular of IPR licensing labelling.
Start Month	03
End Month	13
Milestone(s)	M02.2 Aggregation starts (Month 08)
Deliverable(s)	D02.2 Report on MUSEU aggregation and metadata improvement (Month 13)

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
MUSEU															
Tasks															
T02.1	*Museum sector requirements*			▓	▓	▓	▓	▓	▓						
T02.2	*Metadata improvement*			▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓	▓
Deliverables															
D02.1	*MUSEU data quality museum requirements*								▓						
D02.2	*Report on MUSEU aggregation and metadata improvement*														▓
Milestones															
M02.1	*Release of the museum requirements*								▓						
M02.2	*Aggregation starts*									▓					

3-OpenUp!

Person Months in Task	9

Description of work
Freie Universität Berlin (FUB) will facilitate the standardisation of the metadata provided. Data provision in OpenUp! has quite strictly adhered to the indications given by the Europeana DSI operators, so most if not all OpenUp! records fulfil the criteria set by the guidelines and policies in place. OpenUp! does provide native EDM; OpenUp!'s Natural History Aggregator checks the URL's of providers with every harvest, and there is a continuously working link-checker in place that flags subsequent errors. All multimedia objects have policy-conformant rights-statements, and the records that contain coordinate data are enhanced with place names in the aggregation process. The application of the existing enrichment measures (addition of common names of organisms, synonyms of scientific names, place names etc.) as well as the control of licensing and the technical quality control are included in the current sustainability setup of OpenUp!. Enrichment can be further improved by strengthening the data basis of the web services provided by OpenUp! partners.   All work on data quality improvements will be done in close collaboration with the data partners that provided the data. This includes lots of communication on data quality between all parties including the subcontractor and the data partners.

Subcontracting	
Amount (EUR)	20,000

Access to digital objects in the Europeana DSI is substantially facilitated by enriching the metadata provided with synonyms and multilingual translations. For natural history objects, the names of organisms provide an entry point for the general public and scientists alike. Common names are of particular importance for the general public. During OpenUp!, a web service was created providing common names of organisms when fed with their scientific names. Currently, 470,000 common names in 250 languages from 28 original sources, enhancing the data for 150,000 scientific organism names are included and available for the Europeana DSI. This greatly enhanced access, because the collections themselves do provide only scientific names with their objects. The subcontractor will further strengthen the data basis of the Web service that is used to enhance the metadata in the OpenUp! Aggregator. FUB expects the subcontractor to include at least 20 additional data sources. The actual number of new common names is difficult to estimate; currently the data sources provide an average of more than 16,000 names each. The new sources will probably somewhat smaller, because a special accent will be placed on common names in languages currently under-represented in the Europeana DSI (currently, the bulk of the data are in French, Greek, Italian, Polish, and Portuguese).   The subcontractor must be willing to sustain the service in the long term and needs to be sufficiently ingrained in the natural history community to obtain and judge the quality of new data sources.

Task	T03.1 Integration of additional web sources
Description	Integration of additional web sources into the *CommonNames* web service to facilitate access to natural history content in more languages, e.g. in Dutch, Icelandic, and Chinese
Start Month	04
End Month	12
Milestone(s)	M03.1 Ten additional web sources integrated (Month 08)

	M03.2 Twenty additional web sources integrated (Month 12)
Deliverable(s)	D03.1 Documentation of additional integrated sources (Month 13)

Task	T03.2 Integration of often queried generic terms
Description	Unlike scientists, interested users of Europeana DSI may query natural history objects not by scientifically correct names but by generic terms like 'mouse', and 'chicken'. To satisfy this demand, often queried generic terms will be integrated into the *CommonNames* web service and linked to relevant scientific names. Important terms include for example domestic animals and well-known vermin.
Start Month	04
End Month	12
Milestone(s)	M03.3 Additional generic terms integrated and linked (Month 12)
Deliverable(s)	D03.2 Documentation of additional linked terms (Month 13)

Task	T03.3 Implementation *All-in-one-Service*
Description	The export of common names to enrich OpenUp! metadata will be altered in cooperation with the OpenUp! aggregator. A service will be implemented that allows requesting common names and their sources via a terminal list of scientific names and downloading the results all-in-one shortly after the request. Thus, metadata enrichment is expected to operate more independently.
Start Month	04
End Month	12
Milestone(s)	M03.4 First delivery of common names via *All-in-one-Service* (Month 05) M03.5 Second delivery of common names via *All-in-one-Service* (Month 09)
Deliverable(s)	D03.3 *All-in-one-Service* running stable and with high performance (Month 13)

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
OpenUp!															
Tasks															
T03.1	*Integration of additional web sources*					■	■	■	■	■	■	■	■	■	
T03.2	*Integration of often queried generic terms*					■	■	■	■	■	■	■	■	■	
T03.3	*Implementation All-in-one-Service*					■	■	■	■	■	■	■	■	■	
Deliverables															
D03.1	*Documentation of additional sources*														■
D03.2	*Documentation of additional terms*														■
D03.3	*All-in-one-Service running*						■	■	■	■	■	■	■	■	■
Milestones															
M03.1	*10 additional sources*									■					
M03.2	*20 additional sources*													■	
M03.3	*Additional terms integrated*													■	
M03.4 & M03.5	*Delivery of common names to the OpenUp! Aggregator via All-in-one-Service*						■				■				

4-CARARE

Person Months in Task	1 (2Culture) + 10 (Athena RC)

Description of work
2Culture will work with the CARARE network to plan a roadmap to improvements in data quality, such as the removing barriers to our partners for use of controlled vocabularies and historic place name gazetteers made available as linked open data, and implementing recommendations such as licensing and direct links to digital objects, to make collections compliant to the Europeana Publishing Framework.
2Culture, with Athena RC, will evaluate the enrichment of metadata offered at CARARE aggregator level to communicate the benefits to our partners.
Athena RC will improve existing CARARE data by enriching them in the following ways: enabling SKOS, adding concepts from standard thesauri, adding various types of relations, normalising content. Furthermore, the quality of the provided metadata will be measured, and provide link validation and identity management services.
Athena RC will establish and operate a helpdesk service for the duration of the project, providing metadata support to the CARARE data providers covering all aspects from metadata mapping (e.g. using the MINT tool) to the use of metadata quality / enrichment services in the MORe aggregator infrastructure.

Task	T04.1 Evaluation
Description	This task will evaluate the CARARE data that is already accessible via Europeana (more than 2 million records) and new data to be submitted with the aim of identifying data quality issues, and making a plan of action for improvements. CARARE organised a network meeting which brought data quality issues and Europeana's new publication policy to the attention of network members on 18-19th June 2015. As a follow-up to this meeting: ● 2Culture will circulate a survey to members to find out what content they aim to work on under the Europeana DSI including data and metadata quality issues. ● Athena RC will carry out a review of existing CARARE data with the aim of identifying data quality issues ● Athena RC will review potential enrichment services and innovative technologies that could be used to improve data quality for the domain. ● The results of the survey and reviews will be analysed and an activities plan will be produced.
Start Month	03
End Month	06
Milestone(s)	M04.1 CARARE network meeting (Month 06) M04.2 Survey carried out and results evaluated (Month 06)
Deliverable(s)	D04.1 Activity plan (Month 06)

Task	T04.2 Improve data quality
Description	This task will build on the results of T04.1 and will be responsible for improving metadata and data quality of the aggregated content. Under this task: ● Mappings of existing datasets will be updated by members of the network using MINT, to take developments in both the CARARE schema and the

	implementation of EDM by Europeana.
	● Updated and new datasets and mappings will be ingested to MORe where providers can make use of the services and technologies available to analyse the completeness and otherwise quality assure their metadata.
	● Particular attention will be paid to:
	   ○ Titles and descriptions;
	   ○ The provision of links to digital objects and the availability of digital files to generate previews;
	   ○ The availability of place names and geographic coordinates;
	   ○ subject and period indexing;
	   ○ The coherence of rights statement.
	● The potential to enrich the metadata using the services available through MORe will be evaluated.

Start Month	05
End Month	08
Milestone(s)	M04.3 Improved data quality plan (Month 11)
Deliverable(s)	D04.2 Report on improved content (Month 12)

Task	T04.3 Automatic metadata enrichments
Description	This task will be utilise the microservices available within the MORe repository to validate and enrich the metadata, and will be responsible for executing enrichment plans and delivering enriched content. This task will focus on: ● Normalization incl: language identification, place names, dates & periods; ● Utilization of vocabularies to thematically enrich records; ● Identification of relations between records and contextual resources in Wikipedia; ● Link checking; ● Quality measurement of thumbnails and object links.
Start Month	08
End Month	13
Milestone(s)	M04.4 Metadata enrichment services (Month 13)
Deliverable(s)	D04.3 Report on metadata enrichment (Month 13)

Task	T04.4 Maintenance and update of the CARARE metadata schema
Description	Under this task a working group will be established to maintain and develop the CARARE metadata schema. The working group will: ● Review requests for updating of the schema and make recommendations; ● Maintain the schema documentation and mapping guidelines; ● Maintain the schema XSD; ● Develop an RTF version of the schema; ● Disseminate and promote the CARARE metadata schema.
Start Month	03
End Month	13
Milestone(s)	M04.5 CARARE schema (Month 12)
Deliverable(s)	D04.4 CARARE schema documentation (Month 12)

Task	T04.5 Vocabularies and the domain
Description	This task will build on the results of LoCloud and encourage members of the network to collaborate in the development of domain vocabularies. It will: ● Identify and make available domain related vocabularies; ● Enable collaborative development of domain related vocabularies; ● Enable multilingual developments; ● Provide vocabulary services; ● Enable thematic enrichment of records.
Start Month	08
End Month	13
Deliverable(s)	D04.5 Domain vocabularies and vocabulary services (Month 12)

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
CARARE															
Tasks															
T04.1	*Evaluation*				█	█	█	█							
T04.2	*Improve data quality*						█	█	█	█	█	█	█		
T04.3	*Automatic metadata enrichments*									█	█	█	█	█	█
T04.4	*Maintenance and update of the CARARE metadata schema*				█	█	█	█	█	█	█	█	█	█	█
T04.5	*Vocabularies and the domain*							█	█	█	█	█	█	█	█
Deliverables															
D04.1	*Activity plan*							█							
D04.2	*Report on improved content*													█	
D04.3	*Report on metadata enrichment*														█
D04.4	*CARARE schema documentation*													█	
D04.5	*Domain vocabularies and vocabulary services*													█	
Milestones															
M04.1	*CARARE network meeting*							█							
M04.2	*Survey carried out and results evaluated*							█							
M04.3	*Improved data quality plan*												█		
M04.4	*Metadata enrichment services*														█
M04.5	*CARARE schema*													█	

5-Europeana Fashion

Person Months in Task	5

Description of work
Europeana Fashion will work to improve the aggregated fashion metadata (700,000+ records) according to the Metadata Quality Task Force Recommendations, in particular: ● It will extend the use of the Europeana Fashion multilingual thesaurus for the classification of all the ingested records (maintaining and extending the thesaurus); ● It will develop an authority file of fashion designers to be linked to the actual ingested records; ● It will experiment and apply automatic enrichment techniques (data semantification and image analysis) to add properties such as "colour" to the ingested records (in collaboration with Task 4.5.4); ● It will apply refinement techniques (using the MINT group annotation tool) to further harmonize ingested metadata, especially for names of brands and dates; ● It will add more specific licensing information, encouraging data providers to use Creative Commons licenses and in particular Free Culture licenses. ● It will support data providers in implementing direct links to their provided digital objects (a practice already largely in use in the Europeana Fashion network of data providers).

Task	T05.1 Maintenance and update of the Europeana Fashion Thesaurus
Description	In this task, the existing Europeana Fashion thesaurus will be refined, extended and enriched to meet the needs of the Fashion community at large, also adding an authority file of designer names to be linked with the aggregated datasets. This work will be carried on in collaboration with the thesaurus working group, established during the former Europeana Fashion project. The thesaurus will be published online in a machine-readable format (SKOS), to be re-used openly by all the fashion community.
Start Month	02
End Month	13
Milestone(s)	M05.1 Publishing of the updated version of the Europeana Fashion thesaurus (Month 13)
Deliverable(s)	D05.1 Updated version of the Europeana Fashion Thesaurus (Month 13)

Task	T05.2 Harmonisation, refinement and cleaning of the aggregated fashion data sets
Description	The aggregated fashion datasets (more than 700 thousand records) will be cleaned and harmonised using the group annotation tool developed in the MINT platform. Best practice and recommendations from the Metadata Quality Task Force will be applied during this process. In particular, the refinement work will focus on normalisation of brand and designer names, dates and places using when possible open vocabularies (e.g. VIAF). An increased metadata check will be performed on EDM-mapped metadata prior their publication. Metadata enrichment processes will be applied and validated on all the aggregated datasets (see also T05.3). Also licensing information will be harmonised, trying to encourage the use of Creative Commons licenses. Special attention will be given also to the provision of direct links to digital objects in the metadata records, aiming at reaching 100% of direct links in our metadata records. This task will work in collaboration with Task 1.6.2 "Raising competency levels of data providing partners", in order to disseminate and implement best practices about metadata quality and publication through also the organisation

	of a workshop for data providers. A Basecamp group on content enrichment will be also set-up and used to communicate and offer support to data providers.
Start Month	02
End Month	13
Deliverable(s)	D05.2 Report on fashion data sets cleaning and refinement (Month 13)

Task	T05.3 Automatic metadata enrichment of the aggregated data sets
Description	In this task we will continue the work, started in the former Europeana Fashion project, about automatic enrichment of metadata. In particular we will fine tune and apply image analysis techniques to extract dominant colors from fashion images, and we will fine tune and apply data semantification techniques to enrich the aggregated metadata records extracting relevant terms from the textual descriptions to be added to elements like "Materials", "Techniques" and "Colors". This task will also collaborate with Task 4.5.4 on data enrichment.
Start Month	05
End Month	13
Deliverable(s)	D05.3 Report on semi-automatic metadata enrichment (Month 13)

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
Europeana Fashion															
Tasks															
T05.1	*Maintenance and update of the Europeana Fashion Thesaurus*			■	■	■	■	■	■	■	■	■	■	■	■
T05.2	*Harmonisation, refinement and cleaning of the aggregated fashion datasets*			■	■	■	■	■	■	■	■	■	■	■	■
T05.3	*Automatic metadata enrichment of the aggregated datasets*						■	■	■	■	■	■	■	■	■
Deliverables															
D05.1	*Updated version of the Europeana Fashion Thesaurus*														■
D05.2	*Report on fashion data sets cleaning and refinement*														■
D05.3	*Report on semi-automatic metadata enrichment*														■
Milestones															
M05.1	*Publishing of the updated version of the Europeana Fashion thesaurus*														■

6-EUscreen

Person Months in Task	4

Description of work
The quality of data will be improved through three areas of work. They benefit the overall Europeana infrastructure. In effect, activities in this task will also support the work to reach out to new content providers. ● Europeana Publishing Framework: The aggregators will be informed about the Europeana Publishing Framework and its applicability in their respective working contexts. ● Enrichment: This area of work focuses on enriching the metadata using manual and automatic methods and tools. ● Keyframe extraction.

Task	T06.1 Europeana Publishing Framework
Description	The aggregators will be informed about the Europeana Publishing Framework and its applicability in their respective working contexts.
Start Month	01
End Month	06

Task	T06.2 Enrichment
Description	This area of work focuses on enriching the metadata using manual and automatic methods and tools. In order to improve the quality of the data using enrichment tools, a group-edit functionality will be used that facilitates content providers and data moderators to clean and normalise the data in a user-friendly way. It will give the ability to group items using special search filters and normalise/edit the values using established standards as well as provide suggestions for enrichment from Web resources (e.g. DBpedia) that the users can accept or reject. This work will take the outcomes of *Subtask 4.5.2. Develop EDM* into account.
Start Month	06
End Month	12

Task	T06.3 Keyframe extraction.
Description	Keyframe extraction are added to increase the quality of the Europeana service. An important element of the aggregated data is the visual signifier (i.e. thumbnails) that is sent to the Europeana portal. An automatic screenshot extraction service is developed, securing that all objects contributed to Europeana by EUscreen have keyframes, including correct rights labels.
Start Month	07
End Month	13

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
EUscreen															
Tasks															
T06.1	*Europeana Publishing Framework*	▓	▓	▓	▓	▓	▓	▓							
T06.2	*Enrichment*							▓	▓	▓	▓	▓	▓		
T06.2	*Keyframe extraction*							▓	▓	▓	▓	▓	▓	▓	▓
Milestones															
M06.3	*EUscreen data improvement technical delivery*												▓		

7-Social History Aggregator (HOPE2)

Person Months in Task	8

Description of work
The Social History Aggregator (HOPE2) aims to improve its practices on digital appraisal, materials selected by the new data providers will be reviewed by the core team to define their relevance for the Social History Aggregator and Europeana. The concept of the collection, its completeness and its context is also important to showcase. In case of the existing data providers, some collection descriptions need to be refined, while new partners will go through a stricter quality check prior to considering them for publication. A new policy defines one of the crucial components to provide hierarchical display to digital objects, achieved through elements such as dcterms:isPartOf, edm:isRelatedTo, dc:relation, edm:isNextInSequence.
Semantic enrichment tasks will include applying more multilingual elements, at least for collection descriptions (English and the language of the content) and titles. Translations in 5 European languages for the domain specific themes/subject tags will be added. These tasks will also comprise adding more keywords, subjects, tags connected with skos:Concept class. All HOPE/HOPE2 metadata values using Linked Open Data will be enriched. Metadata should be enriched where possible using Linked Open Data vocabularies to ensure the greater online exposure.
The Social History Aggregator (HOPE2) will highlight the global geographical references to the collections and the extended geographical coverage. The use of the edm:Place class will allow these institutions to submit more specific geospatial information. In addition, ideally, the latitude and longitude of the object will be submitted in the edm:Place class, or edm:TimeSpan class can be also used.
Re-using social history collections is a key aspect of the digital object for both the general and specialist user. Rights statements need to be provided by selecting one of the 13 Europeana rights statements, or providing more clarity on the different licenses and convincing the partners to assign more open licenses. Restrictive statements will be revisited and discussed with existing data providers. The discussion around copyright is especially important when we encourage content providers to add direct links to their digital objects.
Five new data providers will be added (see also Europeana DSI task 1.6.1). Their collections need to be ingested and enriched with themes, language, geospatial elements.

Task	T07.1 Assessing existing metadata in the Social History Aggregator and evaluating possible improvements
Description	Looks at the dataset available on the Social History Portal. Currently, the portal has 183 collections, 45 missing a country element, 20 missing a thematic element, 86 missing a language elements. Further analysis will be carried out on rights values, collection appraisal, direct linking, expanding the tagging tool functionality, reviewing the multilingual metadata values and geospatial information. The result of this analytical work will help renegotiate policies with the current partners and convince them to add the missing values or to invest into the translation of metadata.
Start Month	03
End Month	13
Milestone(s)	M07.1 International Association of Labour History Institutions (IALHI) Annual Meeting to be held in Edinburgh to share the first results (Month 05)
	M07.2 Produce an action plan (Month 05)
	M07 3 Execute action plan and tasks identified (Month 10)

	M07.4 Report back on quality improvement and contribution by current partners (Month 13)
Deliverable(s)	D07.1 Final action plan for social history institutions (Month 06)
	D07.2 Data delivery of improved metadata completed (Month 13)

Task	T07.2 Defining requirements for multilingual access and implementation
Description	Among the semantic enrichment tasks we consider multilingual access one of the most ambitious goals of the project. The current HOPE policy does not prescribe the provision of mandatory bilingual metadata (English and metadata of the content), however for the sake of consistent discovery experience we will highlight the importance for the following elements: collections title and descriptions, item titles, subject terms.The task will also include all HOPE/HOPE2 metadata values using Linked Open Data. Metadata should be enriched where possible using Linked Open Data vocabularies to ensure the greater online exposure.
Start Month	05
End Month	13
Milestone(s)	M07.5 Requirements for multilingual access (Month 05)
	M07.6 Interim report on site visits (Month 09)
	M07.7 Final report on implementation (Month 13)
Deliverable(s)	D07.3 Data delivery of multilingual metadata completed (Month 12)
	D07.4 Applying Linked Open Data vocabularies (Month 13)

Task	T07.3 Defining requirements for geospatial enrichment and implementation
Description	The current social history collections often lack country references while IALHI is a global organisation, we would like to make sure that our international network is well represented from the point of view of our collections as well. The use of the edm:Place class will allow institutions to submit more specific geospatial information. The use of country coverage will enhance discovery in addition to provenance information. Ideally, the latitude and longitude of the object should be submitted in the edm:Place class, or edm:TimeSpan class can be also used. Wherever it is possible we will use automatic geotagging.
Start Month	05
End Month	13
Milestone(s)	M07.8 Requirements for geospatial enrichment and implementation (Month 05)
	M07.9 Interim report on site visits (Month 09)
	M07.10 Final report on the implementation of geotagging (Month 13)
Deliverable(s)	D07.5 Applying geospatial references (Month 12)
	D07.6 Data delivery of geospatial metadata completed (Month 13)

Task	T07.4 Defining requirements for resolving additional access issues: broken links, direct links, formats, technical constraints, licensing
Description	The appearing problems with our current collections could be explained partly by technical constraints edm:isShownBy or edm:isShownAt, lack of persistent ID, technical measures applied to hinder downloads, obsolete formats, non-standard thumbnails. On the other hand the lack of clear license information, missing rights statements and supporting policies on the local collection pages including the Social History Portal represent a conglomerate of problems. Here the enrichment process is more complex, it entails training, consultation, technical work which should be all combined. At the same time the reluctance to rely on more open access principles may conflict with traditional institutional policies, economic exploitation etc. The result of these activities should be monitored through not only having more rights statements with more open

	provisions as metadata in the light of the Euroepana Publishing Framework but communicating more transparent access policies, introducing notice and takedown approaches etc.
Start Month	05
End Month	13
Milestone(s)	M07.11 Revisiting the HOPE IPR Best Practices Recommendations by acknowledging the Europeana Publishing Framework (Month 05) M07.12 Interim report on local IPR situations (Month 09) M07.13 Interim report on local technical problems and possible solutions (Month 09) M07.14 Final report on improved access (Month 13)
Deliverable(s)	D07.7 Updated HOPE WIKI on licensing and IPR including links to local sites (Month 13) D07.8 Data delivery of rights statements (Month 13)

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
Social History Aggregator (HOPE2)															
Tasks															
T07.01	*Assessing existing metadata in the Social History Aggregator and evaluating possible improvements*				■	■	■	■	■	■	■	■	■	■	■
T07.02	*Defining requirements for multilingual access and implementation*						■	■	■	■	■	■	■	■	■
T07.03	*Defining requirements for geospatial enrichment and implementation*						■	■	■	■	■	■	■	■	■
T07.04	*Defining requirements for resolving additional access issues: broken links, direct links, formats, technical constraints, licensing*						■	■	■	■	■	■	■	■	■
Deliverables															
D07.01	*Final action plan for aocial history institutions*						■								
D07.02	*Data delivery of improved metadata completed*														■
D07.03	*Data delivery of multilingual metadata completed*													■	
D07.04	*Applying Linked Open Data vocabularies*														■
D07.05	*Applying geospatial references*													■	
D07.06	*Data delivery of geospatial metadata completed*														■
D07.07	*Updated HOPE WIKI on licensing and IPR including links to local sites*														■
D07.08	*Data delivery of rights statements*														■
Milestones															
M07.01	*International Association of Labour History*						■								

ID	Description
	Institutions (IALHI) Annual Meeting to be held in Edinburgh to share the first results
M07.02	*Produce an action plan*
M07.03	*Execute action plan and tasks identified*
M07.04	*Report back on quality improvement and contribution by current partners*
M07.05	*Requirements for multilingual access*
M07.06	*Interim report on site visits*
M07.07	*Final report on implementation*
M07.08	*Requirements for geospatial enrichment and implementation*
M07.09	*Interim report on site visits*
M07.10	*Final report on the implementation of geotagging*
M07.11	*Revisiting the HOPE IPR Best Practices Recommendations by acknowledging the Europeana Publishing Framework*
M07.12	*Interim report on local IPR situations*
M07.13	*Interim report on local technical problems and possible solutions*
M07.14	*Final report on improved access*

8-The European Library

Person Months in Task	n/a

Description of work
The European Library (TEL) will finalise the migration of its datasets from the ESE format to EDM. The migration to EDM will improve the quality of the data on several important points, such as making all objects provided to Europeana to have persistent identifiers; all objects are labelled with a rights statement from the Europeana rights statements list; all text objects have a language; all metadata records pointing to a digital file or to a html page representing a digital object.

TEL will work on the consistency of rights labels by making sure the rights statements chosen by data providers are consistent with the objects they are applied to. The focus will be made on making sure the objects in the Public Domain are labelled with the Public Domain Mark., TEL will develop validation rules for checking automatically not only by date of publication but also by the digitisation laws that exist per country. This will help to identify objects that belong to the Public Domain but they are not licensed accordingly.

TEL will analyse its collections in Europeana to identify duplicate records and perform deduplication of datasets when needed (e.g. providers with new collections send also duplicate records, that is to say, records included in older sets and published already in TEL and Europeana). TEL would like also to work closer with the data providers in order to determine a new deduplication methodology within the aggregation ecosystem. TEL will provide new recommendations that will be developed further by Europeana to be applicable to all aggregators. These new recommendations will help detecting and preventing duplicate records earlier the aggregation process.

TEL will focus on the normalization of data by increasing the amount of language attributes in the metadata and by normalising the language ISO codes.

TEL will also work on the enrichment of data with controlled vocabularies by increasing the amount of linked data identifiers from AAT, Iconclass, ULAN, TGN used in the metadata. TEL ran a data mining report to identify the amount of identifiers used anywhere in the TEL dataset.

TEL will also work with the data providers to improve the quality of the images (size, resolution, etc) published in Europeana but also to make all digital objects have a thumbnail. TEL identified the number of records with missing thumbnails in Europeana.

Task	T08.1 Moving from ESE to EDM
Description	The European Library will finalise the migration of its datasets from the ESE format to EDM. The migration to EDM will improve the quality of the data on the following points: ● All the objects provided to Europeana have a persistent identifier ● All objects are labelled with a rights statement from the Europeana Rights statements[7] ● All text objects have a language ● All metadata records point to a digital file or to a html page representing a digital object All EDM metadata records will comply with the validation rules described in the EDM documentation.[8]
Start Month	00

---

[7] http://pro.europeana.eu/share-your-data/rights-statement-guidelines/available-rights-statements
[8] http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation

End Month	13
Milestone(s)	M08.1 58 datasets will migrated from ESE to EDM (Month 13)

Task	T08.2 Consistency of rights labels
Description	Make sure the rights statements chosen by data providers are consistent with the objects they are applied to. A specific focus will be put on making sure the objects in the Public Domain are marked Public domain. Validation rules based on the publication date and digitisation laws per country will be put in place.
Start Month	00
End Month	13
Milestone(s)	M08.2 60%-70% of the TEL collections checked, validated with the provider and delivered to Europeana with updated rights statements (Month 13)

Task	T08.3 Deduplication of records
Description	TEL will perform deduplication of datasets when needed (e.g providers with new collections send also duplicate records, that is to say, records included in older sets and published already in TEL and Europeana).
Start Month	00
End Month	13
Milestone(s)	M08.3 30%-50% of the TEL collections checked for duplicates, deduplicated and delivered to Europeana without duplicates (Month 13)

Task	T08.4 Normalisation
Description	Increase the amount of language attributes in the metadata. Normalisation of language ISO codes.
Start Month	00
End Month	13
Milestone(s)	M08.4 30%-50% of the TEL collections checked and delivered to Europeana with increased number of language attributes & normalised ISO codes (Month 13)

Task	T08.5 Vocabulary enrichment
Description	Increase the amount of linked data identifiers from AAT, Iconclass, ULAN, TGN in the data. Identification of number of identifiers used in the TEL dataset done in month 3.
Start Month	00
End Month	13
Milestone(s)	M08.5 30%-50% of the TEL collections checked and delivered to Europeana with enriched vocabulary by increasing the number of linked data identifiers (Month 13)

Task	T08.6 Quality images & thumbnails
Description	TEL will work with its data providers to improve the quality of the images aggregated to Europeana, as defined by the Europeana Publishing Framework. TEL will work with its data providers to increase the amount of thumbnails for all objects provided to Europeana. Identification of TEL records with missing thumbnails in Europeana done in month 3
Start Month	00
End Month	13
Milestone(s)	M08.6 60%-70% of the TEL collections checked and delivered to Europeana with better images and 570,702 TEL records with missing thumbnails from Europeana fixed (with focusing on video and sound material) (Month13)

GANTT chart

Task/Deliverable/Milestone		00	01	02	03	04	05	06	07	08	09	10	11	12	13
The European Library															
Tasks															
T08.1	*Moving from ESE to EDM*	■	■	■	■	■	■	■	■	■	■	■	■	■	■
T08.2	*Consistency of rights labels*	■	■	■	■	■	■	■	■	■	■	■	■	■	■
T08.3	*Deduplication of records*	■	■	■	■	■	■	■	■	■	■	■	■	■	■
T08.4	*Normalisation*	■	■	■	■	■	■	■	■	■	■	■	■	■	■
T08.5	*Vocabulary enrichment*	■	■	■	■	■	■	■	■	■	■	■	■	■	■
T08.6	*Quality images & thumbnails*	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Milestones															
M08.1	*58 datasets will be migrated from ESE to EDM*														■
M08.2	*60-70% of the TEL collections checked, validated with the provider and delivered to Europeana with updated rights statements.*														■
M08.3	*30-50% of the TEL collections checked for duplicates, deduplicated and delivered to Europeana without duplicates*														■
M08.4	*30-50% of the TEL collections checked and delivered to Europeana with increased number of language attributes & normalised ISO codes.*														■
M08.5	*30-50% of the TEL collections checked and delivered to Europeana with enriched vocabulary by increasing the number of linked data identifiers.*														■
M08.6	*60-70% of the TEL collections checked and delivered to Europeana with better images and 570,702 records with missing thumbs fixed*														■