

DELIVERABLE

Project Acronym: Europeana Newspapers

Grant Agreement number: 297380

Project Title: A Gateway to European Newspapers Online

D4.2 Newspaper Aggregation and Indexing Plan

Revision: 4.2

Authors: Alastair Dunning (The European Library/Europeana Foundation)
Markus Muhr (The European Library/Europeana Foundation)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	x
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
0.1	27-02-2012	Markus Muhr	The European Library/Europeana Foundation	Initial version
0.2	27-02-2012	Alastair Dunning	The European Library/Europeana Foundation	Revisions
1.0	June 2013	Alastair Dunning	The European Library/Europeana Foundation	Published (June 2013)
1.1	25-06-2013	Alastair Dunning	The European Library/Europeana Foundation	Adjusted Names of Associated Partners
1.2	31-07-2013	Alastair Dunning	The European Library/Europeana Foundation	Misspelling change
1.3	25-10-2013	Markus Muhr	The European Library/Europeana Foundation	Major Revision
2.0	12-12-2013	Alastair Dunning	The European Library/Europeana Foundation	Rewriting of parts 5 and 6
3.0	10-04-2014	Alastair Dunning	The European Library/Europeana Foundation	Major Revision
3.1	April 2014	Ulrike Zimmermann, Hans-Jörg Lieder	SBB	Review of 3.0
3.2	April 2014	Alastair Dunning	The European Library/Europeana Foundation	Incorporating 3.1 review suggestions
3.3	April 2014	Alastair Dunning	The European Library/Europeana Foundation	Further Revisions
3.4	April 2014	Alastair Dunning	The European Library/Europeana Foundation	Further Revisions
3.5	April 2014	Alastair Dunning	The European Library/Europeana Foundation	Further Revisions
4.0	July 2014	Alastair Dunning	The European Library/Europeana Foundation	Revisions at end of Q2 2014
4.1	31-07-2014	Clemens Neudecker, Hans-Jörg Lieder	SBB	Internal review

4.2	31-07-2014	Sandra Kobel	SBB	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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Table of Contents

1. Executive Summary	5
1.1 Aggregation from Project Partners.....	5
1.2 Aggregation from Associated Partners.....	6
1.3 Forwarding to Europeana	7
2. Introduction	8
3. Developing the Aggregation and Indexing Plan	9
3.1 Europeana Newspaper Content and Associated Partners	9
3.2 OCR Process Plan	9
3.3 Interface Option Paper	9
3.4 Merging Plans to Aggregation and Indexing Plan.....	10
4. Forwarding of Data to Europeana and others.....	11
4. 1 Forwarding Data to Europeana.....	11
4.2 Metadata Quality	11
4.3 Forwarding data to Zeitschriftendatenbank at SBB	12
5. Licencing.....	13
6. Risk Analysis.....	14
7. Aggregation and Indexing Process.....	16
8. Schedule of Aggregation and Indexing Plan.....	17
8.1 July – September 2013 (3rd quarter 2013) - Completed	19
8.2 October – December 2013 (4th quarter 2013) – Completed	19
8.3 January – March 2014 (1st quarter 2014) - Completed.....	19
8.4 April – June 2014 (2nd quarter 2014) -Completed	20
8.5 July – September 2014 (3rd quarter 2014) - Estimated	20
8.6 October – December 2014 (4th quarter 2014) - Estimated	21
8.7 Schedule for Europeana.....	21
Appendix A: Interface Option by Library	23

1. Executive Summary

The newspaper aggregation and indexing plan for Europeana Newspapers provides management for the ingestion process. The content consisting of metadata, full text from OCR results and/or viewing images will be aggregated and ingested into The European Library (TEL), which is also responsible for designing the interface for the project. The metadata will then be forwarded to Europeana and also the Zeitschriftendatenbank at SBB.

This aggregation and indexing plan provides a schedule for ingestion of content from full partners and associated partners. This plan covers the process from getting the data into the central storage on The European Library hosting site, feeding the newspaper interface and finally the delivery of issue-level metadata to Europeana in the Europeana Data Model (EDM).

This plan cuts the ingestion process into small manageable pieces with quarterly deliverables. This allows for flexible changes in the schedule if problems occur as well as gives transparency to the aggregation process. This process covers everything that will be published and accessible via the project website.

1.1 Aggregation from Project Partners

From the 12 project’s library partners, The European Library will aggregate and make available the following via <http://www.theeuropeanlibrary.org/tel4/newspapers> as a core part of the project.

Images	c. 9m scans of digitised newspaper pages (this includes both dynamic and centralised harvesting; for the former TEL does not store the images on its own servers) ¹
Pages of full-text	c. 10m pages of OCR’d text encoded within METS-ALTO format
Issue-Level Records	Issue Level Records relating to 18m pages. This will mean around 1.9m issue records if we assume every issue has 9.5 pages. However, the full number of issue-level records will not be known until TEL has completed aggregation
Article Level Records	Article Level Records related to 2m pages
Title Level Records	From the sample of records ingested already, we expect c.1258 titles related to 18m pages (average of c. 14,300 pages per title) However, it is difficult if not impossible to indicate a reliable number of newspaper titles, mainly because the definition of a newspaper title may differ greatly from country to country

¹ The project as a whole is refining 10m pages; however the libraries are not forwarding all images to TEL. Section 8 goes into more detail on this.

1.2 Aggregation from Associated Partners

For the eleven associated partners, The European Library aims to aggregate and make available the following via <http://www.theeuropeanlibrary.org/tel4/newspapers>.

Associated partners are:

1. National and University Library of Slovenia
2. National Library of Iceland
3. The Royal Library of Belgium
4. National Library of Croatia
5. St. Cyril and Methodius National Library of Bulgaria
6. National Library of Romania
7. National Library of Luxembourg
8. National Library of the Czech Republic
9. National Library of Spain
10. National Library of Portugal
11. National Library of Wales

Note this is in addition to what is required in the Description of Work.

Images	0
Pages of full-text	0
Issue Level Records	Issues level Records relating to a further 19m pages. This will also mean around 2m issue records if we assume every issue has 9.5 pages. However, we will not know the full number of issue-level records until we have done aggregation
Article Level Records	0
Title Level Records	As with full partners, from the sample of records ingested already, we expect c.1329 titles related to 19m pages (average of c. 14,300 pages per title)) However, it is difficult if not impossible to indicate a reliable number of newspaper titles, mainly because the definition of a newspaper title may differ greatly from country to country

The European Library will also host an image server for viewing pages if there is not such one in place at the partner site and if the partner is willing to provide the images to The European Library. The partners' choice for how they wish their content to be displayed is shown in appendix A of this document.

The schedule will be reviewed and updated if necessary every 6 months.

At the end of the project in January 2015, all the content from the libraries listed in the content table will be available through The European Library and delivered in EDM on an issue level to Europeana. Furthermore, full-text content for 10 million pages will be indexed and accessible through the TEL website.

1.3 Forwarding to Europeana

The European Library will convert metadata into the Europeana Data Model (EDM) format and forward it for delivery via Europeana.eu. Following discussion with the Europeana ingestion team, it has been made clear that Europeana.eu does not wish to display page level metadata; instead TEL will provide issue-level metadata that will contain information about the relevant number of articles or pages for each newspaper issue.

Therefore, TEL will forward issue level metadata records relating to 18 million pages to Europeana from the Full Partners in the project. According to the sampling outlined in section 2 this should amount to 1.9 million issue-level records.

As an addition to the core part of the project, TEL aims to forward issue level metadata records relating to 19 million pages to Europeana from the Associated Partners in the project. According to the sampling outlined in section 2 this should amount to a further 2 million issue-level records.

The schedule for delivery of this is outlined in Section 8.

Once issue level records start appearing in Europeana from Q2 2014 onwards, they will be 'flat records' simply referring to metadata about the issue. When the Europeana portal incorporates the ability to display hierarchical records (scheduled for the end of Q3 2014), the metadata ingested by Europeana from TEL will automatically display this hierarchy, for example articles that are present within an issue, provided that such granular information is available in the data. However, it must be noted that this capability for hierarchical records is not part of the Europeana Newspapers Project.

Europeana will also experiment with embedding page images from the TEL server and displaying them on the Europeana.eu site. This will allow users to see the first page of each issue and navigate through that page. Europeana aims to do this for all issues that make up the 9 million images available in TEL.

2. Introduction

This document provides an overview of the aggregation and indexing plan of the project. It provides a realistic schedule for ingesting 18 million pages from newspapers from the full project partners and also for additional aggregation of newspaper content from associated partners. Since the majority of these will be delivered by UIBK and CCS following their OLR / OCR scheduling, this aggregation plan follows their agreed deliverables in the OCR process. These pages constitute a large chunk of data, so that the metadata including OCR full-text and potentially viewing images are being delivered via hard disks.

For the remaining 8 million pages, the images will be made accessible to users via the partners' sites; the issue level metadata for these pages will be transferred to TEL via relevant Internet protocols.

This plan will be split into quarterly scheduling blocks and should be reviewed on a regular basis at the end of every second quarter (6 months).

This document will start with a summary on the objectives and scope of the plan. Secondly, it will outline how the information was gathered by mainly explaining how the survey report guided the development of the aggregation and indexing plan as well as the collaboration with the partners. Thirdly, it will provide a risk analysis. Then, it will outline how the process will be implemented as well as monitored. Finally, a general overview for the individual quarters will be presented.

This newspaper aggregation and indexing plan is vital to the project and should be continuously revisited to allow performance evaluation and risk mitigation.

3. Developing the Aggregation and Indexing Plan

3.1 *Europeana Newspaper Content and Associated Partners*

The initial step for the aggregation plan was to gather all the information from the different partners including newspaper titles, number of issues, number of pages, etc. To get this initial information a survey was created and placed on the project extranet. The results were composed in an excel sheet on a title level. This document also served for the technical partners to select and schedule the OCR/OLR process for different libraries. One major part includes the format of metadata as well as the kind of delivery to The European Library.

The Europeana Newspapers survey identified digitised newspaper content in a number of large European libraries and also led to the identification of eleven associated partners. These were chosen to deliver metadata from their digitised collections and the aggregation has accordingly been included in this document.

3.2 *OCR Process Plan*

Since the aggregation of the newspaper content in The European Library and furthermore indexing the full-text depends on the technical partners UIBK and CCS for more than half of the newspapers, this aggregation plan follows closely the overall processing plan for content refinement. The European Library will get the data directly from the technical partners via hard disks and will further send or upload the data to the University of London Computer Centre, the hosting provider of TEL. The European Library will also make a copy of the hard disks and store those copies on site in The Hague for backup reasons. The format provided by UIBK as well as CCS will be an agreed METS/ALTO profile as defined in Work Package 5 Metadata Best Practice Recommendations. This allows the same treatment for all data provided by the two refinement partners.

3.3 *Interface Option Paper*

The European Library provided the partners with an options paper on how their content should be presented in the TEL website as well as how the viewing images or thumbnails for the search result page should be served.

For all the data aggregated in the Europeana Newspapers Project, metadata will be aggregated and furthermore indexed and is therefore accessible within the interface. Full-text is available for more than half of the 18 million pages and will be indexed as well. Some providers outlined that they have existing full-text available that might go into the interface as well.

However, there were considerable differences on how to serve the viewing images. Basically, there are three options:

- The partner provides an image server that can be accessed by The European Library. No viewing images will be delivered to The European Library by the partner, instead The European Library will request preview images from those image servers and will embed the results into the results shown in the interface;
- The partner has no image server in place, but is willing to provide the viewing images and approves the storage and usage inside an image server maintained by The European Library. In this case the delivered data include, besides metadata and/or full-text viewing,

presentation images. Preview images shown in the interface will be requested from The European Library's own image server.

- The final option is to neither provide the viewing images to The European Library nor give access to an image server on site. In this case, the interface will still show metadata and/or full-text, but only a link back to the specific page on the partner's website is provided instead of viewing images.

The paper with the different options is attached to this document.

3.4 Merging Plans to Aggregation and Indexing Plan

These three resources are the basis for this Aggregation and Indexing Plan. The information was condensed and made centrally available through the project extranet provided by UIBK.

The overall aim of the Aggregation Plan was to balance the workload over the whole project. For this reason, each quarterly chunk includes deliveries from CCS and UIBK which consist of full-text and metadata in a METS/ALTO format. Depending on the option chosen by the partner viewing images can also be part of these deliveries. Furthermore, a quarterly plan includes newspaper content where only metadata will be delivered. Titles for which only metadata will be delivered comes from project partner Austrian National Library as well as from the associated partners.

4. Forwarding of Data to Europeana and others

4.1 Forwarding Data to Europeana

TEL will convert metadata into the Europeana Data Model (EDM) format and forward it to for delivery via Europeana.eu. Following discussion with the Europeana ingestion team, it has been made clear that Europeana.eu does not wish to display page level metadata; instead TEL will provide issue-level metadata that will contain information about the relevant number of articles or pages for each newspaper issue.

Therefore, TEL will forward issue level metadata records relating to 18 million pages to Europeana from the project partners. According to the sampling outlined in section 2 this should amount to 1.9 million issue-level records.

As an addition to the core part of the project, TEL aims to forward issue level metadata records relating to 19 million pages to Europeana from the associated partners in the project. According to the sampling outlined in section 2 this should amount to a further 2 million issue-level records.

The schedule for delivery of this is outlined in Section 8.

Once issue level records start appearing in Europeana from Q2 2014 onwards, they will be 'flat records' simply referring to metadata about the issue. When the Europeana portal incorporates the ability to display hierarchical records (scheduled for the end of Q3 2014), the metadata ingested by Europeana from TEL will automatically display this hierarchy, for examples articles that are present within an issue. However, it must be noted that this capability for hierarchical records is not part of the Europeana Newspapers project.

Europeana will also experiment with embedding page images from the TEL server and displaying them on the Europeana.eu site. This will allow users to see the first page of each issue and navigate through that page. Europeana aims to do this for all 9 million images that are available in TEL.

4.2 Metadata Quality

Each delivery of metadata is sampled and its quality assessed on ingest by Europeana as part of their everyday workflow. If the metadata does not match the EDM guidelines, it will be rejected.

An initial sampling of newspapers metadata converted from native library formats into EDM occurred in Q1 2014. The Europeana ingestion team provided feedback to TEL on the mapping process, and the conversion methodology was updated so that the EDM data was aligned with the general guidelines for EDM and the specific guidelines created in D4.4 *Report on EDM for newspapers*.

As other metadata arrives from other partners in the project, Europeana and TEL will continue to work together to ensure the data is of sufficient quality.

4.3 Forwarding data to Zeitschriftendatenbank at SBB

The European Library will also forward Dublin Core metadata about each issue to be curated by the Zeitschriftendatenbank at SBB as part of Deliverable D4.3 *Report on European Registry for Digital Newspapers Online* in M30. This will follow the same schedule for delivery as the forwarding of metadata to the Europeana Foundation (EF).

5. Licencing

As part of the Consortium Agreement, all full partners agreed to release their metadata as CC0.² As the decision to include metadata from the associated partners was taken after the project, TEL is working on the best way to ensure that their metadata is released as CC0 as well.

As part of another piece of work, outside the scope of this project, TEL is currently working with all its member libraries (which include all associated partners in the Europeana Newspapers project) to mark metadata as CC0 as part of its TEL Partnership Agreement. Europeana has agreed that this is sufficient to get the data marked as CC0; indeed this is Europeana's preferred method of handling the issue.³

For other content (e.g. full text and images), the libraries are placing rights statements on the metadata that governs the use of the content.

TEL has also requested that Europeana provide further guidance in terms of the rights statements for full text and images. The recent changes to the available rights statements (<http://pro.europeana.eu/available-rights-statements>) have caused confusion over which statements should be applied for historic newspapers, in particular the usage of the 'Out of copyright – non-commercial re-use (OOC-NC)' statement.

To help with this, Europeana is currently updating its publication policy and will publish an advocacy programme on public domain and opening up collections. Once this has been published, TEL will work with Europeana to ensure that the partners in the project are using the correct rights statements, communicating with projects through the rest of the project, and using the planned Newspapers policy workshop on 29 September 2014 in London as a mechanism for discussing the issue in more depth.

² https://sp.uibk.ac.at/sites/eu-news/coordination/SiteAssets/D1.1.2_ENP_Consortium%20Agreement.pdf
Article 9.1.

³ Verbal correspondence with Gina van den Linden, Europeana Foundation.

6. Risk Analysis

Risk in delays is omnipresent, so we need to have an agile approach to address changes in delivery schedules from partners. For this reason, we will adapt this document over time. Delivery cycles are quarterly to allow us some freedom on changing scheduling. This document will be revisited after every six months. Furthermore, the updates should continue to be in line with changes in the OCR scheduling, since these directly affect the aggregation status.

Risk	Outcome	Possibility	Impact	Score	Mitigation
Aggregation of content via OCR partners is delayed due to delays from partner libraries	Project misses core target of 17m pages	4	5	20	Keeping in contact with OCR partners and pressuring libraries where necessary; rescheduling content from different sources
Aggregation of full-text delayed because of quantity of data	Ingestion delayed	5	3	15	Ingest high volume content (e.g. from KB) in smaller doses
Aggregation of content from OCR partners is delayed due to processing problems (e.g. OCR partner has unforeseen technical problems and OCR process stands still for days)	Project misses core target of 18m pages	1	4	4	Keeping in frequent contact with OCR partners; rescheduling content from different sources
Aggregation of content from OCR partners is delayed due to Transfer problems (e.g. hard disks get lost on the way)	Project misses core target of 18m pages	2	1	2	Close monitoring of delivery
Availability of stable links at partners site is delayed	Metadata does not appear on TEL or Europeana; project misses target	3	1	3	Identify struggling libraries and assist them; source other content or host all content at TEL (last option currently being used)
Aggregation of metadata from partners is delayed due to The European Library's lack of support for certain metadata formats or transfer protocols	Metadata does not appear on TEL or Europeana; project misses target	2	2	4	Work with partners to improve if possible in short term, If from Associate Partners, accept
Aggregation of metadata from Associate Partners is delayed due to lack of funding or required skills	Associate partner metadata not in TEL or Europeana	4	1	4	Accept; not a core part of project

for these					
Delivery to Europeana is delayed due to Insufficient EDM support	Metadata does not appear in Europeana; project misses target	2	4	8	TEL has worked with Europeana to improve EDM mapping
Delivery to Europeana is delayed due to Blocked ingestion process (e.g. too many collections in pipeline)	Metadata does not appear in Europeana; project misses target	3	4	12	Are continuously working with Europeana ingestion team to balance schedule
Delivery to Europeana is delayed due to wrong rights labelling statements by libraries	Metadata does not appear in Europeana; project misses target	2	4	8	Are working with libraries to work to get correct rights statements
Delivery to Europeana is delayed due to lack of response on sharing CC0 (Associate Partners only)	Metadata does not appear in Europeana	3	1	3	Accept; not a core part of project

7. Aggregation and Indexing Process

There are four key activities that run sequentially. First, aggregation of data by The European Library, which consists partly of hard disks (UIBK, CSS) and partly of metadata, harvested via the internet from partners. The data on the hard disks need then to be sent further to London to be put onto the servers of The European Library (hosted by University of London Computer Centre) or will be directly uploaded from the office in The Hague in case of small deliveries.

The second phase includes ingesting the metadata into the services at TEL and will then make them accessible from the portal (www.theeuropeanlibrary.org/).

The third phase involves indexing the full-text in a separate search index optimised for this purpose and uploading if necessary the viewing images to the image server. After this step, the interface will be filled with the data and the newspapers can be retrieved from there.

Fourth, on an issue level the objects are delivered to Europeana in EDM and will then be retrievable from www.europeana.eu at the latest one month after delivery. This last step is repeated to provide the objects on an issue level to the Zeitschriftendatenbank (ZDB) at SBB.

8. Schedule of Aggregation and Indexing Plan

Below is a summary of the number of pages, full-text pages, and images to be aggregated by The European Library during the project. TEL will be able to calculate the number of issues received as it harvests the data.

The aggregation of newspaper full-text, metadata and/or images started in the 3rd quarter of 2013 (July to September). The other quarterly aggregations are (or have been) planned for October 2013, January 2014, April 2014, July 2014 and October 2014. The last quarter of the project should aggregate and index only final outstanding refinement results.

The project aims to make at least c. 37.000.000 pages accessible on or through both TEL and Europeana – this relates to content from both full (about 18 million pages) and associated partners (about 19 million pages).⁴ Records representing each issue within titles will be delivered to Europeana and will be accessible through the Europeana portal (www.europeana.eu). All the pages will be accessible on an individual level either through full-text and metadata search in the interface developed by TEL. Pages are normally made available on the TEL interface between four and six weeks after aggregation.

Around 10 million pages of full text and 4 million images will be aggregated by TEL. Note that some of the images delivered by the TEL interface will not be stored by TEL but remotely called when a user searches for a specific title. Around 9 million images will therefore appear on the TEL interface, but only less than half of those will actually be aggregated and held centrally by TEL.

The project as a whole is refining about 10 million newspaper pages; about 9 million images will be delivered via TEL. The National Library of Turkey images are behind a paywall and they are not being released as open access. The SBB is only delivering 1.1 million of the foreseen 1.7 million images. The project is currently working with the National Library of Luxembourg, Dr Friedrich Tessman Library and others to bring in extra 0.6 million images. At the moment of writing, the full workflow for the Dr Friedrich Tessman Library is still being scheduled, and is not included in the current figures. The National Library of Luxembourg figures remain classified as associated partners. TEL has also organised a meeting with the National Library of France (BnF) to discuss how they can reach the 3 million count originally planned – as only 2.4 million are currently scheduled to come available via the refinement process.

The tables in the following chapters show the aggregation plan per quarter. The table columns are the acronym for the provider, the country of the provider, the distribution way (via CCS, UIBK or directly from partner for non-refined material), the target number of pages in the Description of Work done, and the number of pages, full-text and images to be aggregated and indexed in the specific quarter.

The aggregation plan follows the common ingestion plan of The European Library, so that the ingestion process will benefit from the experience of the team there. The last quarter in the aggregation plan is smaller in size to give TEL some buffer to finalise the aggregation process and to make sure all relevant content is accessible through the TEL interface and from Europeana.

⁴ The precise number of issues is unknown until all metadata is received.

At the time of writing, TEL is still awaiting the first delivery from CCS, which has seen some delays due to more extensive quality control of the refinement results by libraries. The content arriving from SBB via UIBK has also been delayed. TEL has aggregated metadata and full-text related to the BnF collections but is still involved in discussions with the BnF about how best to access their images servers so that the collections can be served remotely. Once an agreement is reached with BnF, around 2.4 million more pages that have already been refined in the project can also become available in the portal.

Acronyms used in the following tables for OCR partners:

CCS – Content Conversion Specialists

UIBK – University of Innsbruck

Acronyms used in the following tables for partner libraries:

BnF - Bibliotheque nationale de France / National Library France

KB - Koninklijke Bibliotheek / National Library of the Netherlands

LFT Landesbibliothek Dr. Friedrich Teßmann / Teßmann Library

NLE - Eesti Rahvusraamatukogu / Estonian National Library

NLF – Kansalliskirjasto / National Library of Finland

NLL - Latvijas Nacionālā Bibliotēka / National Library of Latvia

NLP - Biblioteka Narodowa / National Library of Poland

NLT - Milli Kutuphane Baskanligi / National Library of Turkey

ONB - Österreichische Nationalbibliothek / Austrian National Library

SBB - Staatsbibliothek zu Berlin / Berlin State Library

SUBHH - Staats- und Universitätsbibliothek Hamburg / State and University Library

UB - Univerzitet u Beogradu / University Library of Belgrade

Acronyms used in the following tables for Associated Partners:

NLW - National Library of Wales

NLB - St. Cyril and Methodius National Library / National Library of Bulgaria

NLC - National Library of Czech Republic

NLH - National and University Library in Zagreb

KBR - Koninklijke Bibliotheek van België / Bibliothèque royale de Belgique

NUK - Narodna in univerzitetna knjižnica / National and University Library of Slovenia

NLPR - National Library of Portugal

NRO - National Library of Romania

NLS – National Library of Serbia

NLSlo – National Library of Slovakia

LI - Landsbókasafn Íslands - Háskólabókasafn / National and University Library of Iceland

BNE – National Library of Spain

BNL - Bibliothèque nationale de Luxembourg / National Library of Luxembourg

8.1 July – September 2013 (3rd quarter 2013) - Completed

Provider	Country	Via	DoW	# pages	# full-text	# images
ONB	AUT	UIBK	7057772	1089641	1089641	0
NLF	FIN	UIBK	114000	40579	40579	40579

Aggregated this quarter = 1,130,220 pages

Full project partner total = 1,130,220 pages

Associated partner total = 0

8.2 October – December 2013 (4th quarter 2013) – Completed

Provider	Country	VIA	DoW	# pages	# full-text	# images
LFT	ITA	UIBK	840082	851935	851935	851935
BnF	FRA	CCS	3000000	-	-	-
BnF	FRA	UIBK	3000000	1000000	1000000	0
NLL	LTV	UIBK	460781	454563	454563	454563
ONB*	AUT	ONB	7057772	139722	139722	-
LI	ICE	LI	0	4112602	0	0
BNE	ESP	BNE	0	5831826	0	0

Aggregated this quarter = 12,390,648 pages

Full project partner total = 3,576,440 pages

Associated partner total = 9,944,428 pages

8.3 January – March 2014 (1st quarter 2014) - Completed

Provider	Country	Via	DoW	# pages	# full-text	#images
BnF	FRA	CCS	3000000	0	0	0
NLE	EST	UIBK	594657	94698	94698	94701
NLE	EST	CCS	594657	0	0	0
SUBHH	GER	UIBK	1545000	400000	400000	400000
SBB	GER	UIBK	1710000	0	0	0
UB	SRB	UIBK	387700	242208	242208	242208
SUBHH	GER	CCS	1545000	0	0	0
NLP	POL	UIBK	99106	83648	83648	83648
UB	SER	UIBK	387700	145592	145592	145592

Aggregated this quarter = 966,146 pages
 Full project partner total = 4,542,586 pages
 Associated partner total = 9,944,428 pages

8.4 April – June 2014 (2nd quarter 2014) - Completed

Provider	Country	Via	DoW	# pages	# full-text	#images
BnF	FRA	CCS	3,000,000	200,000	200,000	-
NLE	EST	CCS	594,657	-	-	-
SUBHH	GER	CCS	1,545,000	-	-	-
KB	NET	KB	2,000,000	-	-	-
ONB	AUT	UIBK	7,057,772	75,204	75,204	-
SBB	GER	UIBK	1,710,000	121,261	121,261	121,261

8.5 July – September 2014 (3rd quarter 2014) - Estimated

Provider	Country	Via	DoW	# pages	# full-text	#images
BnF	FRA	CCS	3,000,000	450,000	450,000	-
NLE	EST	CCS	594,657	50,000	50,000	50,000
SUBHH	GER	CCS	1,545,000	100,000	100,000	100,000
KB	NED	KB	2,000,000	1,921,946	1,921,946	-
NLE	EST	CCS	594,657	300,000	300,000	300,000
SUBHH	GER	CCS	1,545,000	200,000	200,000	200,000
SBB	GER	CCS	1,710,000	10,000	10,000	10,000
ONB	AUT	ONB	7,057,772	195,433	195,433	-
KBR	BEL	KBR	-	3,500,000	-	-
NUK	SLOV	NUK	-	-	-	-
NLPR	PRT	NLPR	-	400,000	-	-
NRO	ROM	NRO	-	442,398	-	-
NLT	NLT	NLT	400,000	400,000	-	-
NLSlo	SLO	SLO	-	-	-	-
NLS	SER	NLS	-	-	-	-

NLW	GBR	NLW	-	1,100,000	-	-
NLB	BUL	NLB	-	12,000	-	-
NLC	CZE	NLC	-	2,760,000	-	-
NLH	CRO	NLH	-	300,000	-	-
BNL	LUX	BNL	-	620,000	106,562	106,562

8.6 October – December 2014 (4th quarter 2014) - Estimated⁵

Provider	Country	Via	DoW	# pages	# full-text	#images
NLF	FIN	CCS	114,000	46,428	46,428	46,428
BnF	FRA	CCS	3,000,000	197,761	197,761	-
NLE	EST	CCS	594,657	129,962	129,962	129,962
SUBHH	GER	CCS	1,545,000	180,000	180,000	180,000
NLF	FIN	CCS	114,000	45,000	45,000	45,000
BnF	FRA	UIBK	3,000,000	570,000	570,000	-
SUBHH	GER	UIBK	1,545,000	665,000	665,000	665,000
SBB	GER	UIBK	1,710,000	600,000	600,000	600,000
ONB	AUT	ONB	7,057,772	5,557,772	-	-

8.7 Schedule for Europeana

Beginning from April 2014, each batch of data incoming to The European Library will be delivered to Europeana at the following dates. The first data is scheduled to appear in Europeana in May 2014.

Schedule for when TEL will forward data to Europeana ingestion team

	Q3, 2013	Q4, 2013	Q1, 2014	Q2, 2014	Q3, 2014	Q4, 2014
Full pages	1,130,296	2,446,220	1,012,548	396,465	3,627,379	7,991,923

⁵ This precise schedule for this quarter still requires further conversation with UIBK and CCS to confirm final numbers arriving.

Full issues	118,979	257,497	106,584	41,733	381,829	841,255
Associate pages	-	9,944,428	-	4,792,000	4,342,398	-
Associate issues	-	1,046,782	-	504,421	457,095	-

Data is then published in the Europeana portal in the following month, e.g. data delivered to Europeana in June 2014 (at the end of Q2), will be published in July 2014 (Q3). This also assumes that all licencing conditions are met as mentioned in Section 5.

- a) Full project partners have assigned the rights statements to their metadata, and
- b) Associated partners sign up to the TEL partnership agreement to allow for their data to be released as CC0.

Appendix A: Interface Option by Library

Based on figures from start of project

Library	Country	Metadata	Full-text	Images	Serving images	Interface Choice
UB	SRB	408.181	408.181	408.181	Images served by TEL	Option 1
SBB	GER	248.200	248.200	248.200	Images served by TEL	Option 1
ONB	AUT	6.781.332	1.090.308	0	Images served by ONB	Option 1
NLP	POL	83.648	83.648	83.648	Images served by TEL	Option 1
NLL	LTV	460.781	460.781	460.781	Images served by TEL	Option 1
NLF	FIN	132.093	132.093	132.093	Images served by NLF	Option 1
NLE	EST	594.663	594.663	594.663	Images served by TEL	Option 1
LFT	GER	857.485	857.485	857.485	Images served by TEL	Option 1
KB	NET	1.921.946	1.921.946	0	Images served by KB	Option 1
SUBHH	GER	2.216.200	2.216.200	2.216.200	Images served by TEL	Option 1
BnF	FRA	2.388.488	2.388.488	0	Images served by BnF	Option 2
NLT	TRK	8.990	0	0	No images in interface, referral to webpage at NLT	Option 4