



# MILESTONE

**Project Acronym:** Europeana Cloud

**Grant Agreement number:** 325091

**Project Title:** Europeana Cloud: Unlocking Europe's Research via The Cloud

## MS20: Update report on metadata ingestion

(January 2015)

Revision	final
Date of submission	29 January 2015
Author(s)	Marian Lefferts, CERL Adina Ciocoiu, Europeana Foundation
Dissemination Level	Public

Project co-funded by the European Commission within the ICT Policy Support Programme

## REVISION HISTORY AND STATEMENT OF ORIGINALITY

### Revision History

Revision No.	Date	Author	Organisation	Description
1 <sup>st</sup> draft	26 January 2015	M. Lefferts	CERL	Documenting progress report discussed with Adina Ciocoiu
2 <sup>nd</sup> draft	28 January 2015	M. Lefferts	CERL	Processed additional comments by Adina Ciocoiu and Alastair Dunning
Final	29 January 2015	M. Lefferts	CERL	

### 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This milestone contains original unpublished work except where clearly indicated otherwise.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

1. Updated Metadata Ingestion Plan .....	5
2. Datasets processed in project to date .....	7

## Updated Metadata Ingestion Plan

The tables below contain an overview of the data set listed in the Description of Work, and not yet processed.

*Table 1 Overview ingestion process*

Quarter	Action	Number of records
Q1 2015	8 collections	78,559
Q2 2015	4 collections	At least 4,870
Q3 2015	4 collections	At least 3,500
Q4 2015	last deliveries to Europeana, deadline 21 November 2015	
	<b>Total</b>	

*Table 2 Ingestion plan for Q1 2015*

Provider	Country	Collections	Records	Digital Objects	DoW
Questa.Soft	<b>Germany</b>	Philology, Linguistics, Psychology, Education	20.000	0	52,800 (The C.E.E.O.L. database contains 160.000 articles in PDF format in Philology, linguistics, psychology, education)
Questa.Soft	<b>Germany</b>	Political and Social Studies, Economy, Law	10.000	0	60,800 (Political and social studies, economy, law)
Questa.Soft	<b>Germany</b>	History, Judaic Studies, Philosophy, Religion, fine and performing Arts	10.000	0	46,400 (The C.E.E.O.L. database contains 46,400 History, Judaic studies, philosophy, religion, fine and performing arts)
Open Knowledge Foundation	<b>Germany</b>	Will ingest directly into the Cloud			
HAZU	<b>Croatia</b>	The remaining 7 textual collections on Croatian arts, history and science, including Documents for	12.559		

		the History of Croatian Literature and Miscellany on Folk Life and Customs			
The University and National Library of Debrecen	<b>Hungary</b>	19th Century Periodicals from the Eastern Hungarian Region	10.000	950	6500
The University and National Library of Debrecen	<b>Hungary</b>	18th and early 19th Century Hungarian Periodicals	10.000	1.200	5000
The University and National Library of Debrecen	<b>Hungary</b>	Playbills of the Debrecen Theatre 1862-1896	6.000	6.000	6.500 records

Table 3 Ingestion plan for Q2 2015

NL Denmark	<b>Denmark</b>	- Archive of Danish Literature - Articles from the Tidsskrift.dk			
Library and Information Center, University of Patras	<b>Greece</b>	Pasitheo	1.327	24.298	1.400 (Journals representing more than 25,000 pages)
VU University Amsterdam Library	<b>Netherlands</b>	Gedigitaliseerde boeken en tijdschriften	3.543	120.000	2.493 records

Table 4 Ingestion plan for Q3 2015

Tilburg University	<b>Netherlands</b>	Tilburg University Dataverse	150	150	100 (100 research datasets and supplementary materials)
Kansallis arkisto	<b>Finland</b>	Europeana Heraldica	3.200	3.200	3200 (2200 coats of arms, 1000 seals)
CESSDA	<b>Sweden</b>				25,000 research datasets
Cinecittà	<b>Italy</b>			100	Film reels



## Datasets processed in project to date

The project partners' metadata sets are being processed by The European Library as per the Metadata Ingestion Plan (available via Basecamp and Europeana Pro).

Particularly the institutions aggregating metadata from a variety of sources, and then passing it on to Europeana are experiencing great difficulties in complying with the demand for full and consistent rights labelling. This why data processed by The European Library is not always included in Europeana.

Often they would wish to use the rights label 'unknown'. Data thus labelled is not currently accepted in Europeana. It is explored whether this ruling could be reviewed.

All records processed during this process and held in the database of The European Library will be available in the Europeana Cloud infrastructure and are therefore in principle available for the Europeana Research facility for scholars, academics and researchers that is developed in this project.

		TEL	Europeana	Comments
Q2 2013	OAPEN	4,730	2,010	Contradiction between dc:rights and edm:rights in the source data causes the discrepancy in numbers. Regular updates
Q3 2013	BASE	758,248	0	Over 36,000 strings with (incomplete/wrong/absent) rights statements that cannot be checked as they are aggregated from many contributors. Proposal to adopt 'unknown' is not acceptable in Europeana portal.
	DANS	29,339	2,819	Restricted rights records with no previews to digital objects were excluded from Europeana Portal
	VU Amsterdam	4,778	4,575	
Q4 2013	University of Edinburgh	13,544	12,478	Regular updates
	Université libre de Bruxelles	420	257	Contain records with 'unknown' rights statement, as the date of death of the author could not be determined. These are excluded from the Europeana Portal. Regular updates
Q1 2014	Dialnet	4,512	4,086	Not all records have links to digital objects, so these were invalidated by Europeana
	NL Wales	36	36	
	Questa	0	0	3 sets in acceptance, but dc:language is missing, plus issues with isShownAt / ShownBy. We need new export,

				but response from provider is lacking
	Tilburg			Need to clarify their IPR statement. Postponed to Q3 2015
	VU Amsterdam	3,156	2,967	Regular updates
	BSB München	1.030.768	1,025,191	BSB prepared the data in EDM, which was time consuming. Finally an issue around the correct EDM labelling of multiple copies for the same imprint, resolved in autumn 2014, caused further delays. Some issues e.g. missing date of expiration for Out of Copyright material are outstanding.
Q2 2014	Sibiu	187	187	
	Patras	(1,327 in preparation)		Postponed until Q1 2015
	NL Denmark	336,743	336,741	
	National Library of Technology in Prague	228	228	
	UCL	16,960	575	Disagreement on correct rights label for Bentham collection, and a smaller set with restricted access, both not accepted in Europeana Portal
Q3 2014	Croatian Academy of Sciences and Arts	6,405	6,405	Further sets scheduled for Q4 2014 and in 2015
	Patras	22,576	22,576	
	VU Amsterdam	92	92	
Q4 2014	CESSDA			Change in management. Rescheduled for Q3 2015
	Sibiu	53	53	
	Croatian Academy of Sciences and Arts	5,011	In preparation	
	University of Leuven	20,500	In preparation	
	<b>Totals</b>	<b>2,255,130</b>	<b>1,421,276</b>	

Outstanding in 2015 are at least 86,929 records<sup>1</sup>, which would bring the total of metadata records processed in this project 2,342,059. The target for the project was 2.5 million, and

<sup>1</sup> See table I above



we aim in the final year of the project to make up the shortfall of some 158k records (e.g. from the extra contributions by the NL Denmark and the Open Knowledge Foundation).