DELIVERABLE

Project Acronym:	**Europeana Cloud**
Grant Agreement number:	**325091**
Project Title:	**Europeana Cloud: Unlocking Europe's Research via The Cloud**

Deliverable D4.1 – Metadata Ingestion Plan

Revision: Final

Authors:

Markus Muhr (The European Library)
Chiara Latronico (The European Library)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
PU	Public	

Revision History

Revision	Date	Author	Organisation	Description
1	15.05.13	Chiara Latronico	TEL	Initial timetables
2	17.05.13	Markus Muhr	TEL	Initial draft
3	21.05.13	Chiara Latronico	TEL	Update draft
4	22.05.13	Marian Lefferts	CERL	Revised draft
5	22.05.13	Chiara Latronico	TEL	Updated tables
6	22.05.13	Marian Lefferts	CERL	Light editing
7	30.05.13	Marian Lefferts	CERL	Processed comments Reviewer
FINAL	30.05.13	Chiara Latronico	TEL	Updated tables

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Executive Summary

The Metadata Ingestion Plan for Europeana Cloud supports the process of managing the workflows for aggregating and ingesting metadata through The European Library into Europeana.

This plan presents the ingestion process in general and a realistic timeframe for ingestion of all metadata delivered by the content providers of the project. The European Library will serve as the main aggregator in this project and will deliver aggregated metadata datasets to Europeana for ingestion. This plan builds on successful and efficient workflows, aggregation and ingestion processes, established during the Europeana Libraries project (which ended in December 2012).

2.4 million new records will be made available through the portals of The European Library and Europeana by the end of the project in 2015. The basis for this plan was a questionnaire filled out by all partners (see Appendix A). The answers, especially on the readiness of the data, guided the timetabling of the datasets. The ingestion plan is broken down into manageable portions of quarterly planned datasets for ingestion. This plan will be reviewed and updated every six months, rescheduling datasets that were not delivered to timetable. Quieter periods (such as Quarter 3 in 2014 and Quarter 4 in 2015) ensure that any delayed datasets can easily be accommodated and do not negatively impact the ingestion timetable for the other datasets.

This ingestion plan will start with an outline on the scope of the ingestion task. It will then outline the process how The European Library has gathered the information guiding this ingestion plan from the partners. Then it will highlight the aggregation and ingestion process in The European Library as well as how this can be monitored. Risk assessment and mitigation will be introduced in this section as well. In the following section the quarterly plan for the ingestion is introduced. The last section will wrap up this plan and highlight next steps in the ingestion plan. The Metadata Ingestion Plan is an important tool for performance evaluation and risk management.

Executive Summary 1
1. Introduction 2
2. Development of the Metadata Ingestion Plan 3
3. Aggregation and Ingestion Process 4
4. Risk Analysis 6
5. Introducing the Ingestion Plan 8
6. Conclusion 16
7. Appendix A: Questionnaire 17
8. Appendix B: SugarCRM 20
9. Appendix C: Tools in the Ingestion Process 22

1. Introduction

The Metadata Ingestion Plan for Europeana Cloud establishes a realistic schedule for the aggregation and ingestion of the metadata from partners. The European Library will utilize their established processes, for harvesting the data, ingestion into the live portal of The European Library, acceptance of data by partners and, finally, data delivery to Europeana. The schedule of the ingestion process has been created using the results of a questionnaire circulated to those project partners committed to contributing metadata. The plan will be reviewed, and where necessary updated every six months, and is flexible enough to adapt to unforeseeable problems with contributing partners, The European Library or Europeana. The Metadata Ingestion Plan has scheduled more intense ingestion activities in the first half of the project, as in the second year of the project the focus will shift towards the ingestion of content. By the end of the project 2.4+ million new metadata records and 5 million items of high quality digital content will be available for reuse by researchers.

2. Development of the Metadata Ingestion Plan

The Metadata Ingestion Plan was created in cooperation with the Europeana Cloud content partners. The information was collected from the partners via a questionnaire from the participating institutions. The questionnaire was based on general questions that The European Library asks all new providers, but tailored to the requirements of the project. It was sent out in March 2013, with a deadline of early May 2013.

The questionnaire asked about general information of the partner institutions, specifics about their datasets and, especially interesting for this plan, the time of availability of their data. The questions of the questionnaire and a screenshot of the Google form that was utilized for this purpose can be found in Appendix A.

The results of the questionnaire were validated and then transferred to the Customer Relationship Management tool (SugarCRM) in use at The European Library. Using SugarCRM, a first timeline of the ingestion process was established. SugarCRM allows us to easily keep track of the ingestion process as automatically generated reports immediately show the status of ingested and planned datasets per quarter. A screenshot of these reports is provided in Appendix B.

The initial timeline was refined based on expertise of the operations team in The European Library, with the aim to balance the workload and especially datasets that are anticipated to be more difficult to aggregate and ingest with those that are expected to be easier to process. A dataset is easier to process when it can be ingested directly in the system in the original metadata format. Metadata standards with direct support are: MARC21, MarcXchange, Marcxml, Unimarc, ESE, EDM, OAI_DC. A dataset is more difficult to process when the original metadata format (e.g. METS, MODS, TEI) needs to be mapped to one of the formats above or if it needs major data cleaning.

In June and July 2013, the first ingestion clinics will be held to show all participating institutions the aggregation and ingestion workflow performed by The European Library, using the specific example of datasets contributed for this project. A follow-up will be organised in conjunction with the Annual Meeting organised by The European Library (Amsterdam, 17-18 September 2013), and further Ingestion Clinics will be held at regular intervals during the project to allow partners to address their specific needs and problems for the ingestion process. Each of these clinics should give us feedback on the ingestion process and might lead to updates of the Ingestion Plan.

3. Aggregation and Ingestion Process

The aggregation and further ingestion into The European Library starts in quarter 2 of 2013 with two initial collections. These two initial collections will also serve as examples of the ingestion process for the first ingestion clinic in June/July 2013, consisting of three Google Hang-out sessions for 8-9 project partners each, on 19 June, 21 June, and 3 July.

This document representing the Metadata Ingestion plan will be updated regularly. Informal updates can occur after each quarter and official updates will be done every 6 months during the runtime of the project. The overview of the ingestion timetable can be seen in Table 1. A split into quarters achieves a good trade-off between flexibility and efficiency of the plan.

Table 1 Overview ingestion process

Quarter	Action	Number of records
Q1 2013	Planning	
Q2 2013	2 collections	3.414
Q3 2013	13 collections	903.160
Q4 2013	18 collections	1.016.428
Q1 2014	8 collections	44.650
Q2 2014	9 collections	36.864
Q3 2014	1 collection	3.543
Q4 2014	4 collections	20.530
Q1 2015	3 collections	26.000
Q2 2015	0 collections	0
Q3 2015	2 collections	3.200
Q4 2015	last deliveries to Europeana	
	Total	**2.057.789**

As can be seen from the table, the plan aims to process most of the metadata in the first half of the project, to allow staff to be reallocated to the content ingestion planned for the third year of the project. Most of the ingestion will be take place in quarters 3 and 4 of 2013. Quarters 3 of 2014 and 2 and 4 of 2015 act as a buffer, during which time delayed content may be processed without jeopardising the overall timetable

Note that this plan is missing two providers: Cinecitta', KB Denmark, who did not fill in the questionnaire. KB Denmark is itself an aggregator for Europeana and has indicated that it wishes to make alternative arrangements for contributing its (c. 400k items) to Europeana. The dataset pledged by Cinecitta' is relatively small (c. 100 newsreels) and can be accommodated in the general schedule without upsetting the general timetable. This means that the target of ingesting 2.4 million metadata records will comfortably be met, well before the end of the project (January 2016).

The general aggregation and indexing process of The European Library consists of the following steps:

1. Content ingestion questionnaire
2. Scheduling of ingestion
3. Datasets ready for harvesting
4. Create case in CRM: case # sent to provider

5. Harvesting metadata (Tool Repox see Appendix C)
6. Enhance metadata (VIAF, Geonames, MACS, etc.) (Tool UIM see Appendix C)
7. Indexing in acceptance portal (http://www.tel.ulcc.ac.uk/acceptance/see Appendix C)
8. E-mail to provider to accept dataset
9. Live index = live portal (http://www.theeuropeanlibrary.org/)
10. Delivery to Europeana
11. Enhancing and publishing in Europeana

Note that actions 1-4 have already been completed for the data from Europeana Cloud in the preparation of this ingestion plan.

The Metadata Ingestion Plan has two major activities: aggregation and ingestion into The European Library portal and delivery to Europeana (after partners have accepted dataset presentation in The European Library). Since the period between delivery to and ingestion into Europeana may be around one month, the ingested datasets may be accessible in the Europeana portal only in the quarter following the quarter indicated in Table 1 above.

4. Risk Analysis

Risk assessment and mitigation is of the highest importance.

The risks that have been identified so far are:

- Aggregation of metadata is delayed due to problems on the provider site (hardware failures, delay in availability of data, metadata format problems, unforeseen legal problems etc.)
    - Impact: Dataset cannot be aggregated and ingested by The European Library and will therefore not appear in The European Library or Europeana
    - Score: Low to medium probability, very severe if it is not just a temporary delay
    - Mitigation: If delays are only temporarily, reschedule dataset at a later stage in the project. We should keep good contacts with the partners to foresee these problems early. Problems with metadata should be recognized early by evaluating provided samples in time.
- The European Library is not able to aggregate and ingest data or deliver data to Europeana.
    - Impact: This would yield to a very high impact if this problem is severe and cannot be solved in a short period of time as The European Library is the single point to get the data from the provider and deliver them to Europeana.
    - Score: Due to stable and robust ingestion processes and deliveries to Europeana the probability should be very low, but if a big hardware failure or problems with data amounts arise, it can have a very severe effect on the project.
    - Mitigation: This should be communicated immediately to the project board and worst-case scenario might include switching aggregator. In case of a temporary outage, the plan should be adapted and changes must be communicated to the project board.
- Europeana cannot process data in time (hardware failures, system upgrades, etc.)
    - Impact: The content is not available in Europeana.
    - Score: Especially in the beginning of the project, this might be a higher risk than usually for the reason that Europeana is going through a major system change including switching to Europeana Data Model (EDM) and using Unified Ingestion Manager (UIM) for ingestion. The portal as well went through a major system upgrade and might suffer from some initial bugs. This can be severe if the problems are still there after the first year.
    - Mitigation: The European Library should keep a close cooperation with the Europeana Ingestion Team and timely address problems with deliveries in EDM.

To ensure that we do not experience delays on the ingestion timetable the Ingestion Team of the European Library maintains regular communications with all data providers to ensure a timely delivery of data.

To further mitigate risks in delays in the ingestion process due to unknown metadata standards etc. we ask all partners to provide us a data sample of their data which is especially important for data formats that are not directly supported by TEL (DC and MARC variants preferred). The Ingestion Team analyses sample sets of data to ensure that major

problems with datasets are recognised early, and a dataset is rescheduled if more pre-processing (e.g. format transformations) is required.

Slippage on the overall timetable is minimised by provision of buffers. The regular review of the Ingestion Plan, plus its division in stages of three months each allow for a close monitoring of progress, and timely counter measures, when required. Regular meetings between the ingestion team and the WP leader should guarantee a good supervision of the ingestion process.

All risks identified above are closely monitored by The European Library. During the runtime of the project problems that arise with the ingestion process will be reported to the work package leader. If problems are bigger than expected, these are escalated to Project Manager, and, if necessary, to the Project Board. If the risk probability for any of the risks above significantly increases or a major risk like unavailability of The European Library infrastructure occurs, this is reported immediately to the Project Manager, through the WP leader.

5. Introducing the Ingestion Plan

Ingestion is planned on a quarterly basis starting with April 2013 (Q2 2013) and ending on December 2015 (Q4 2015). The ingestion of metadata has the highest concentration in 2013 as data is already available and project ingestion at this stage is compatible with the general workload for The European Library. Note that Quarters 3 of 2014 and 2 and 4 of 2015 act as a buffer, during which time delayed content may be processed without jeopardising the overall timetable.

Table 2 Ingestion plan for Q2 2013

Provider	Country	Collections	Records	Digital Objects	DoW
OAPEN Foundation	**Netherlands**	OAPEN	2.000	0	2.000 (Approx. 1000 peer reviewed monographs, from more than 30 European academic publishers)
OAPEN Foundation	**Netherlands**	DOAB	1.414	0	
		2 collections	**3.414**	**0**	

Table 3 Ingestion plan for Q3 2013

Provider	Country	Collections	Records	Digital Objects	DoW
Bielefeld UL	**Germany**	BASE Humanities & Social Sciences	866.245	0	Approximately 1,500,000 (including material from outside European repositories); About 1,559,000 records (out of about 37,800,000 records in total), relating to Humanities & Social Sciences. Restricting this result to European repositories only, we have currently about 733,000 records
DANS	**Netherlands**	Social Sciences	2.500	0	?[3]
DANS	**Netherlands**	Humanities	25.000	0	?
DANS	**Netherlands**	Socio-Cultural Sciences	5.000	0	?

[3] When a quotation mark (?) is found in the DoW column, it means that this information was not available when the DoW was prepared.

VU University Amsterdam Library	**Netherlands**	Fotocollectie Historisch Documentatiecentrum voor het Nederlands Protestantisme (1800-heden)	749	749	6,000 records[4]
VU University Amsterdam Library	**Netherlands**	Amsterdam Law Forum	180	180	180 records
VU University Amsterdam Library	**Netherlands**	Kunstlicht	82	700	?
VU University Amsterdam Library	**Netherlands**	Portretten	1.456	1.510	1.456 records
VU University Amsterdam Library	**Netherlands**	Historische kaarten	1.297	4.821	1.266 records
VU University Amsterdam Library	**Netherlands**	Medieval Manuscripts	9	3.223	9 records
VU University Amsterdam Library	**Netherlands**	Brieven	295	1.369	295 records
VU University Amsterdam Library	**Netherlands**	Moderne handschriften	18	3.107	18 records
VU University Amsterdam Library	**Netherlands**	Collectie van Noordwijk	329	1.967	300 records
		13 collections	**903.160**	**17.626**	

Table 4 Ingestion plan for Q4 2013

Provider	Country	Collections	Records	Digital Objects	DoW
Bavarian State Library	**Germany**	Works Digitized by Bavarian State Library and other Bavarian Libraries	1.000.000	0	350k [2]
HAZU	**Croatia**	17 textual collections on Croatian arts, history and	8.775	322.227	8,775 (17 textual collections comprising of 322,227

[4] VU Amsterdam University Library and the Bavarian Libraries have been asked to clarify the difference between their offerings in the DoW and their offerings in the questionnaire.

		science, including Documents for the History of Croatian Literature and Miscellany on Folk Life and Customs			documents)
HAZU	**Croatia**	Images drawn from Croatian Gliptotheque and Museum of Architecture on art and history from 13th century onwards	830	830	830 (12 visual collections comprising 830 images)
ULB	**Belgium**	Pirenne & Laurent	191	191	191 (8,200 pages from books and articles)
ULB	**Belgium**	Literary Journals	10	10	8 (29,900 pages; complete collections of 8 journals)
ULB	**Belgium**	19th-20th Century Books	35	35	36 (Currently 161,300 pages from 27 books and 9 journal titles)
ULB	**Belgium**	Pierre Gilbert	35	35	35 (2,100 pages from books and articles, comprising articles, reviews, chapters, biographies and reports)
ULB	**Belgium**	University Press of ULB	157	157	157 (36,800 pages, around 76 books, 81 journals)
University of Edinburgh	**United Kingdom**	Hill and Adamson	701	701	5,000+ (Specialized collections Architectural Drawings=286 images; Ars Anatomica=550 images; Charting The Nation=407 owned by Edinburgh; Hill & Adamson=700; Laing=1,330; religious manuscripts=2,000+; Salvesen=175)
University of Edinburgh	**United Kingdom**	Architectural Drawings	285	285	
University of Edinburgh	**United Kingdom**	David Laing	1.330	1.330	
University of Edinburgh	**United Kingdom**	Christian Salvesen	175	175	
University of Edinburgh	**United Kingdom**	Oriental Manuscripts	686	686	
University of Edinburgh	**United Kingdom**	Western Medieval Manuscripts	1.882	1.882	
University	**United**	School of Scottish	44	44	2,500+ (Folklore

of Edinburgh	**Kingdom**	Studies			collections Fine Art=500 high quality images; Walter Scott=1,070 images; Scottish Studies=44; University of Edinburgh=500; Thomson-Walker=135; other Gallimaufry)
University of Edinburgh	**United Kingdom**	New College	203	203	
University of Edinburgh	**United Kingdom**	Incunabula	18	18	
University of Edinburgh	**United Kingdom**	Walter Scott	1.071	1.071	
		18 collections	**1.016.428**	**471.214**	

Table 5 Ingestion plan for Q1 2014

Provider	Country	Collections	Records	Digital Objects	DoW
Dialnet	**Spain**	Artículos de revistas editadas por el Instituto de Estudios Riojanos (Articles from journals edited by IER)	2.000	0	2,115 (3,182 journal Articles; 3072 with full-text; 721 with resume); 748 (3,182 journal articles; 3,072 with full-text; 721 with resume); 618 (3,182 journal articles; 3,072 with full-text; 721 with resume)
Dialnet	**Spain**	Artículos de revistas editados por Universidad de La Rioja	2.200	0	
National Library of Wales	**United Kingdom**	Historical Welsh Journals	200	180.000	200 (180,000 pages of digitised journals = digital images and text. 200 title level records and 180,000 page level records.)
Questa.Soft	**Germany**	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	**Germany**	Political and Social Studies, Economy, Law	10.000	0	60,800 (Political and social studies, economy, law)
Questa.Soft	**Germany**	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)
Tilburg University	**Netherlands**	Tilburg University Dataverse	150	150	100 (100 research datasets and supplementary materials)
VU University Amsterdam Library	**Netherlands**	Historische VU films en video's	100	100	100 records
		8 collections	**44,650**	**180,250**	

Table 6 Ingestion plan for Q2 2014

Provider	Country	Collections	Records	Digital Objects	DoW
"Lucian Blaga" University of Sibiu	**Romania**	Brukenthal Museum - Calendars and Journals	100	3.500	220 (10,000 pages scanned from the 19th century German calendars and journals, journals, old and rare books = 19 books and 46 separate items; 6500 pages scanned, about 20 books descriptions, 200 historical images
National Technical Library	**Czech Republic**	Historical Maps	106	4.773	?
Library and Information Center, University of Patras	**Greece**	Pasithee	1.327	24.298	1.400 (Journals representing more than 25000 pages)
Library and Information Center, University of Patras	**Greece**	Pleias	22.596	65.858	23,000 (82,000 journal Articles=280,000 pages, 100 books=10,000 pages)
National Technical Library	**Czech Republic**	Historical Monographs	84	14.895	159 (9 journals approx. 50,000 pages 150 monographs approx. 25,000 pages)
National Technical Library	**Czech Republic**	Historical Periodicals	157	36.283	
University College London	**United Kingdom**	Orwell	320	320	442 (Digitised archival material: 221, 18 literary notebooks and diaries, 202 photos, 1 letter.)

University College London	**United Kingdom**	CAVA	1.284	1.284	722 (722 digital objects representing 250 hours of video.)
University College London	**United Kingdom**	Bentham	10.890	10.890	3.183
		9 collections	**36.864**	**162.101**	

Table 7 Ingestion plan for Q3 2014

Provider	Country	Collections	Records	Digital Objects	DoW
VU University Amsterdam Library	**Netherlands**	Gedigitaliseerde boeken en tijdschriften	3.543	120.000	2.493 records
		1 collection	**3.543**	**120.000**	

Table 8 Ingestion plan for Q4 2014

Provider	Country	Collections	Records	Digital Objects	DoW
"Lucian Blaga" University of Sibiu	**Romania**	Old Romanian Journals and Books	30	6.500	20 (10,000 pages scanned from the 19th century German calendars and journals, journals, old and rare books = 19 books and 46 separate items; 3500 pages scanned, about 20 titles descriptions)
University of Leuven	**Belgium**	Doctoral Dissertations - Old University	9.000	9.000	2.500 books
University of Leuven	**Belgium**	Pamphlets	1.500	1.500	1.350 images
University of Leuven	**Belgium**	Postcards	10.000	10.000	1,000 images + 2000 items
		4 collections	**20.530**	**27.000**	

Table 9 Ingestion plan for Q1 2015

Provider	Country	Collections	Records	Digital Objects	DoW
The University and National Library of Debrecen	**Hungary**	19th Century Periodicals from the Eastern Hungarian Region	10.000	950	6500
The University and National Library of Debrecen	**Hungary**	18th and early 19th Century Hungarian Periodicals	10.000	1.200	5000
The University and National Library of Debrecen	**Hungary**	Playbills of the Debrecen Theatre 1862-1896	6.000	6.000	6.500 records
		3 collections	**26.000**	**8.150**	

Table 10 Ingestion plan for Q2 2015 - BUFFER

Provider	Country	Collections	Records	Digital Objects	DoW
		0 collections	**0**		

Table 11 Ingestion plan for Q3 2015

Provider	Country	Collections	Records	Digital Objects	DoW
Kansallisarkisto	**Finland**	Europeana Heraldica	3.200	3.200	3200 (2200 coats of arms, 1000 seals)
Open Knowledge Foundation	**Germany**	not supplied [5]	not supplied	not supplied	
		2 collections	**3.200**	**3.200**	

[5] The Open Knowledge Foundation will be asked for clarifications nearer the time their data will be processed.

Table 12 Ingestion plan for Q4 2015 - BUFFER

Provider	Country	Collections	Records	Digital Objects	DoW

6. Conclusion

The Metadata Ingestion Plan is a key tool for monitoring the aggregation and ingestion of new metadata records in Europeana Cloud. The process describes how the data from the content providers in the project is aggregated and ingested into The European Library and further delivered to Europeana.

The Metadata Ingestion Plan was created in cooperation with the content providers, The European Library and Europeana. It outlines how the information was gathered via a questionnaire and subsequently transformed into this plan. It highlights the procedure The European Library will apply to bring the data from providers to the Europeana portal. The plan stretches the ingestion process starting with Q2 2013 and ending in Q4 2015. At the end of the project 2.4 millions records will be available in the Europeana portal.

After creating this plan The European Library will start aggregating and ingesting the first datasets which will serve as examples for the first Ingestion Clinic to showcase the process and get feedback from the partners on the ingestion process.

7. Appendix A: Questionnaire

Entry Page

1. ***Personal Information (about the person filling in the web-form):***

- Mr/Mrs/Miss;
- First name;
- Family name;
- Title;
- Email address;
- Skype address.

Second Page

2. ***Information about Organization:***

- Organization name;
- Country;
- Website;
- Type of institution.

Third Page

3. ***Harvesting Details:***

- Which protocol will be used to transfer data to The European Library? [OAI-PMH, File, Z39.50, other];
- Harvesting time and dates preferences; If yes, which would be the preferred time? [out-of-office hours, weekends, other; If yes, which ones would be the preferred dates? If yes, for which collection(s)?
- How often your collection(s) will need to be updated? [weekly, monthly, quarterly, etc.]; If yes, for which collection(s)?

Fourth Page

4. ***Information about Dataset(s)/Collection(s):***

- Number of dataset(s)/collection(s) to be ingested;
- Collection(s) name(s);
- Number of records to be expected per collection(s);
- Number of digital objects to be expected per collection(s);

- Contact person(s) per collection(s): first name, family name, email, phone, Skype, area of expertise [Editorial: for collection description, Technical: for collection ingestion]

5. Information about Metadata:

- Metadata standard(s) available to describe objects (Marc formats are strongly recommended) [Marc21, Unimarc, ESE, EDM, METS, MODS, OAI_DC, TEI, other];
- How many formats are available per dataset?
- Are the metadata ready? If yes, for which collection(s)? If not, when will they be ready?
- Type of digital objects per collection(s) [TEXT, IMAGE, AUDIO, VIDEO].

6. Information about Content:

- Will content be delivered in addition to metadata? If yes, for which collection(s)? If yes, in which format(s)?
- Has the content been digitized? If yes, for which collection(s)? If not, when will it be?

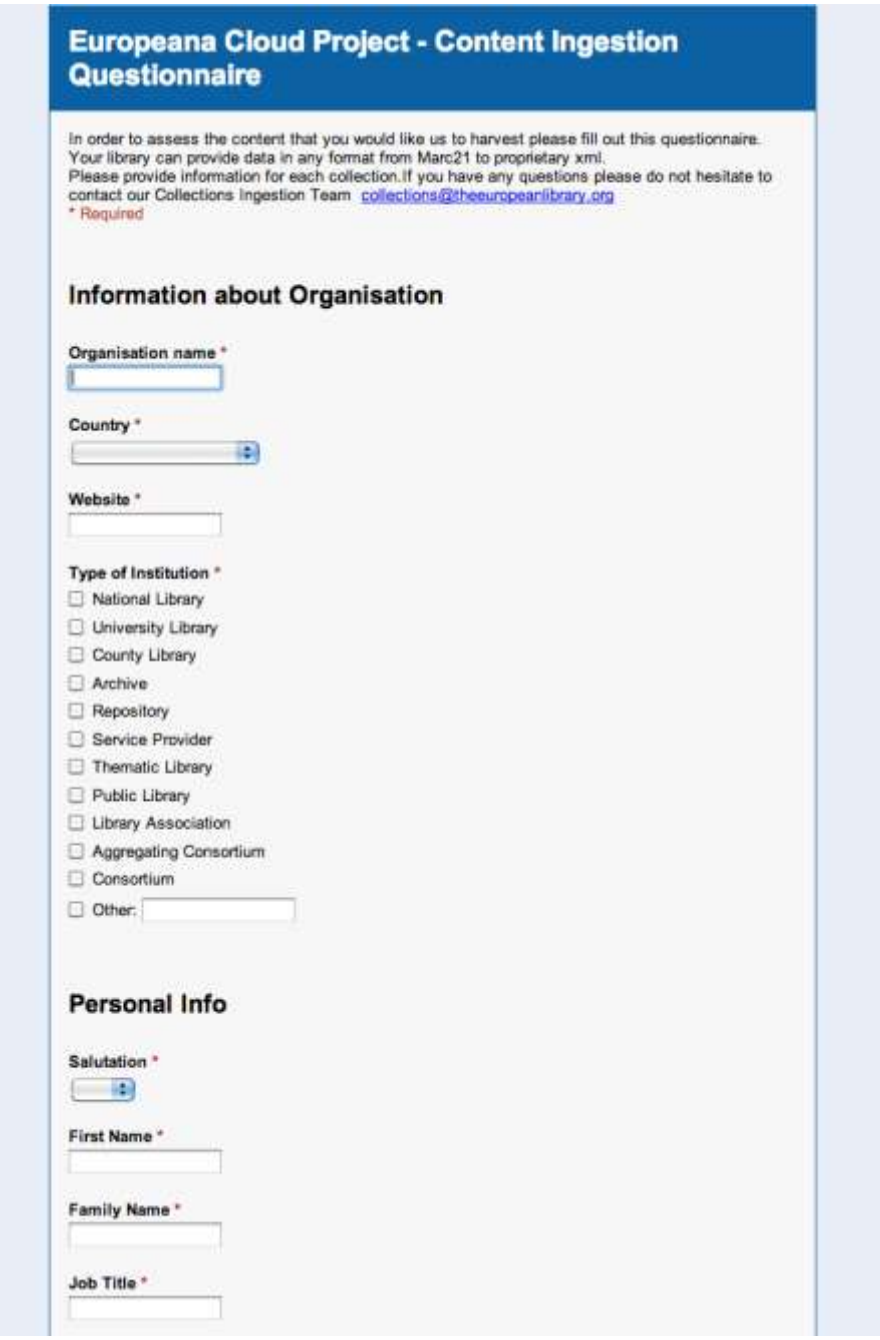
7. Information about Authority:

- Will authority files be delivered? If yes, for which collection(s)?
- Are controlled vocabularies utilized? If yes, which kind? [Classification, Thesauri, Subject Headings, other]; If yes, for which collection(s)?
- Will full-text be delivered? If yes for which collection(s)?

Fifth/Last Page of the online questionnaire:

If you made a mistake we will be happy to help!

Send an email to collections@theeuropeanlibrary.org or chiara.latronico@kb.nl



Europeana Cloud Project - Content Ingestion Questionnaire

In order to assess the content that you would like us to harvest please fill out this questionnaire.
Your library can provide data in any format from Marc21 to proprietary xml.
Please provide information for each collection.If you have any questions please do not hesitate to contact our Collections Ingestion Team collections@theeuropeanlibrary.org
* Required

Information about Organisation

Organisation name *

Country *

Website *

Type of Institution *

- ☐ National Library
- ☐ University Library
- ☐ County Library
- ☐ Archive
- ☐ Repository
- ☐ Service Provider
- ☐ Thematic Library
- ☐ Public Library
- ☐ Library Association
- ☐ Aggregating Consortium
- ☐ Consortium
- ☐ Other:

Personal Info

Salutation *

First Name *

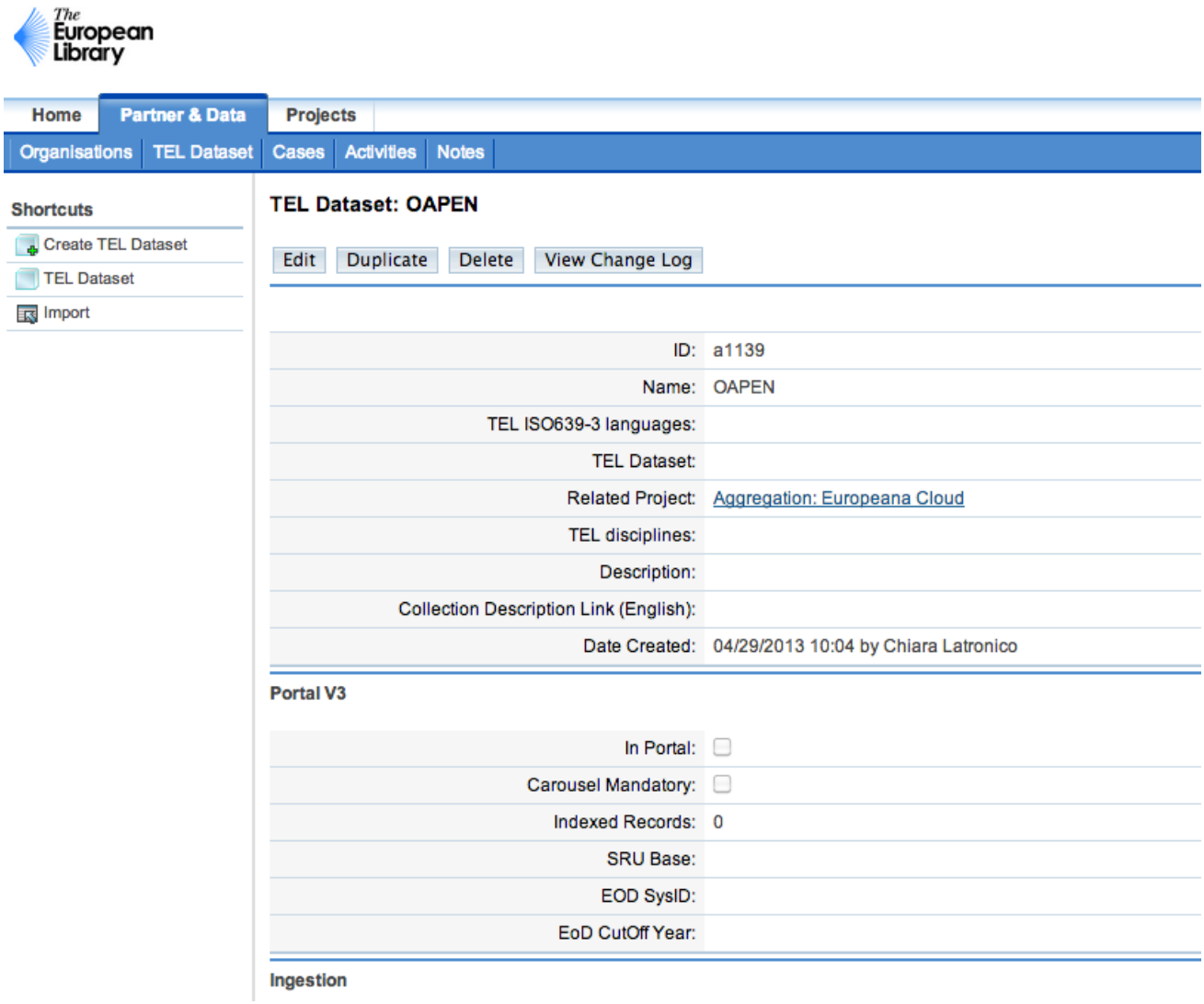
Family Name *

Job Title *

8. Appendix B: SugarCRM

User interface

SugarCRM is a Customer Relationship Management tool in which all information about partners, datasets, right statements, dates of harvesting and ingestion, delivery to Europeana, etc. is stored. It automatically creates reports and synchronizes data with other systems in use at The European Library, such as the Portal, UIM and Repox.



Reports for ingestion process

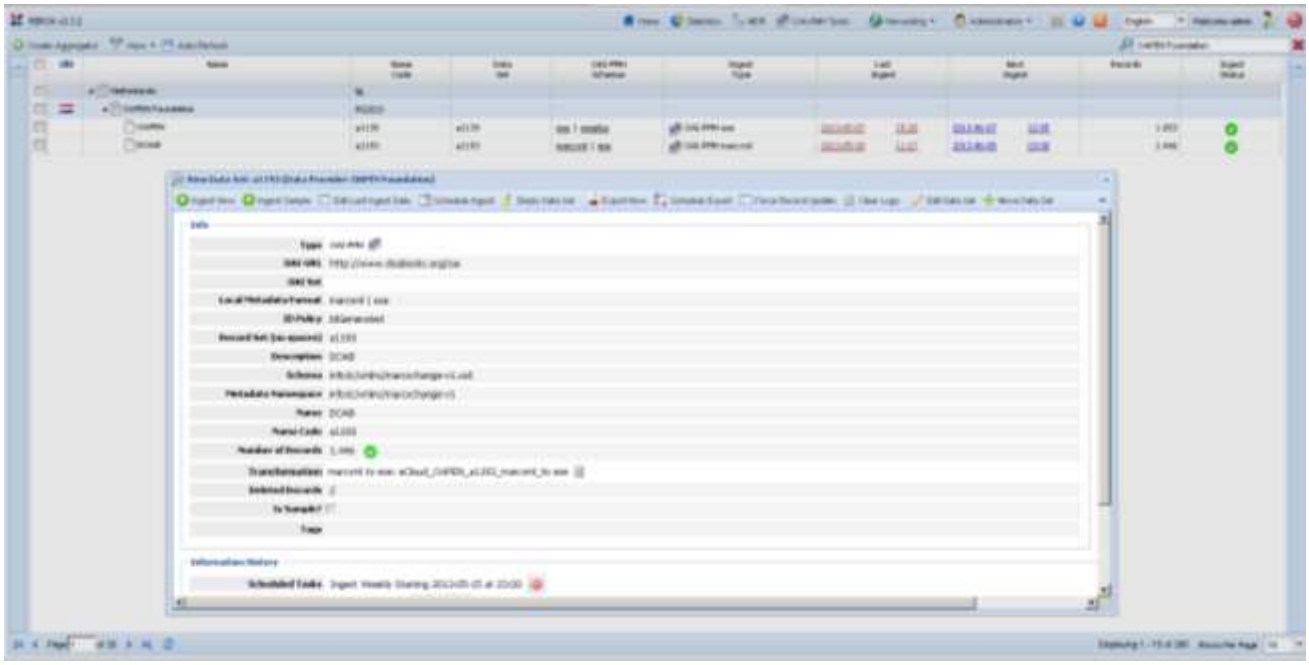
Ingestion Plan:Aggregation: Europeana Cloud

The Hague May 21, 2013

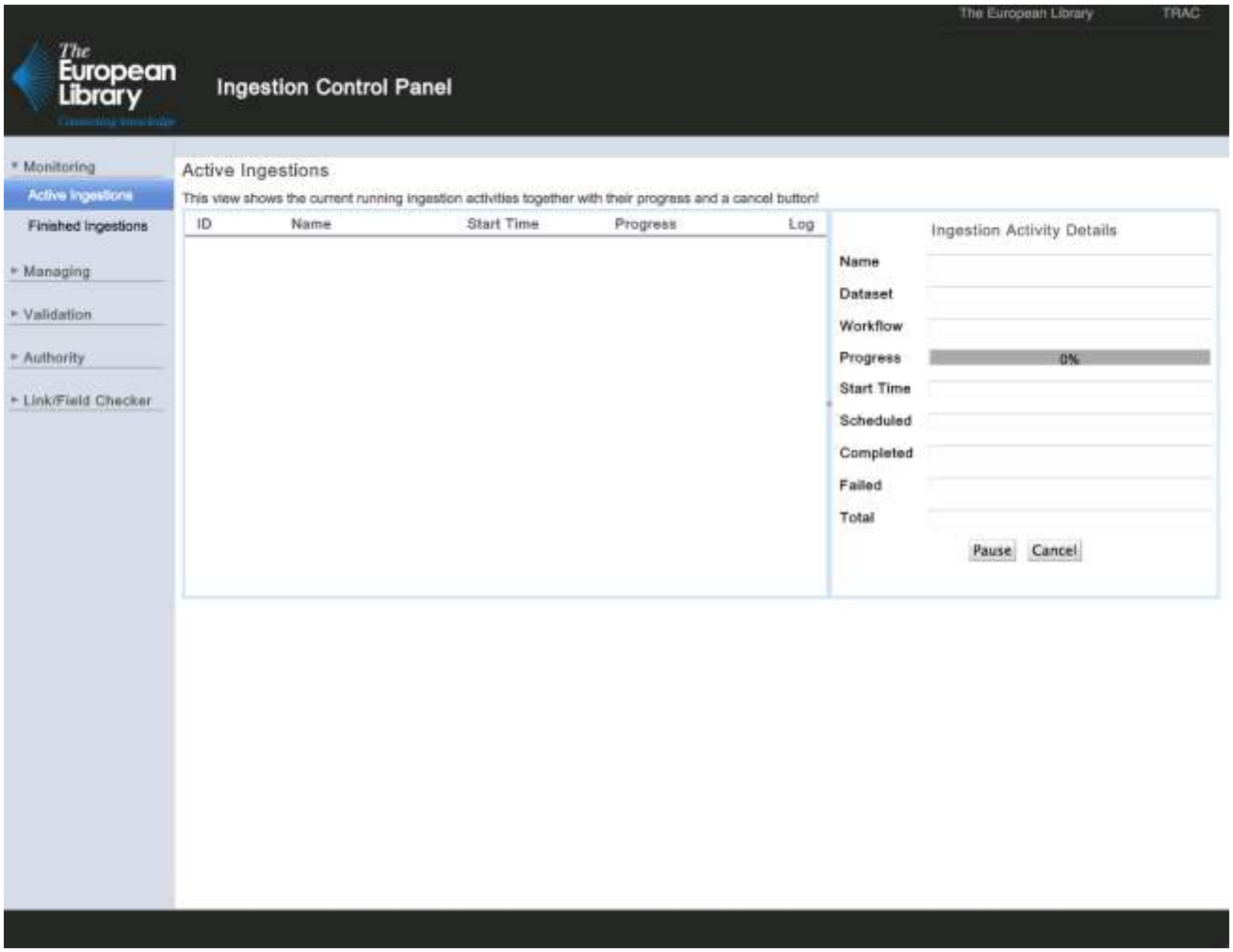
Aggregation: Europeana Cloud Task: Q1 2013 - TEL Processing (EC)			2013-01-01	2013-03-31
Q1 2013 - TEL Processing (EC)				
Aggregation: Europeana Cloud Task: Q2 2013 - TEL Processing (EC)			**2013-04-01**	**2013-06-30**
OAPEN Foundation	Status	Expected	Harvested	Portal
a1139: OAPEN (in ESE)	intaccept	2,000	1,764	0
a1193: DOAB (in MARC21)	acceptance	1,414	1,415	0
OAPEN Foundation		3,414	3,179	0
Q2 2013 - TEL Processing (EC)		**3,414**	**3,179**	**0**
Aggregation: Europeana Cloud Task: Q3 2013 - TEL Processing (EC)			**2013-07-01**	**2013-09-30**
Bielefeld UL	Status	Expected	Harvested	Portal
a1141: BASE Humanities & Social Siences (in OAIDC)	preparation	866,245		
Bielefeld UL		866,245		
DANS	Status	Expected	Harvested	Portal
a1162: Social Sciences (in OAIDC)	preparation	2,500		
a1161: Humanities (in OAIDC)	preparation	25,000		
a1163: Socio-Cultural Sciences (in OAIDC)	preparation	5,000		
DANS		32,500		
VU University Amsterdam Library	Status	Expected	Harvested	Portal
a1152: Fotocollectie Historisch Documentatiecentrum voor het Nederlands Protestantisme (1800-heden) (in OAIDC)	preparation	749		
a1160: Amsterdam Law Forum (in MARC21)	preparation	180		
a1159: Kunstlicht (in OAIDC)	preparation	82		
a1155: Portretten (in OAIDC)	preparation	1,456		
a1154: Historische kaarten (in OAIDC)	preparation	1,297		
a1150: Medieval Manuscripts (in OAIDC)	preparation	9		
a1157: Brieven (in OAIDC)	preparation	295		
a1151: Moderne handschriften (in OAIDC)	preparation	18		
a1156: Collectie van Noordwijk (in OAIDC)	preparation	329		
VU University Amsterdam Library		4,415		
Q3 2013 - TEL Processing (EC)		**903,160**		

9. Appendix C: Tools in the Ingestion Process

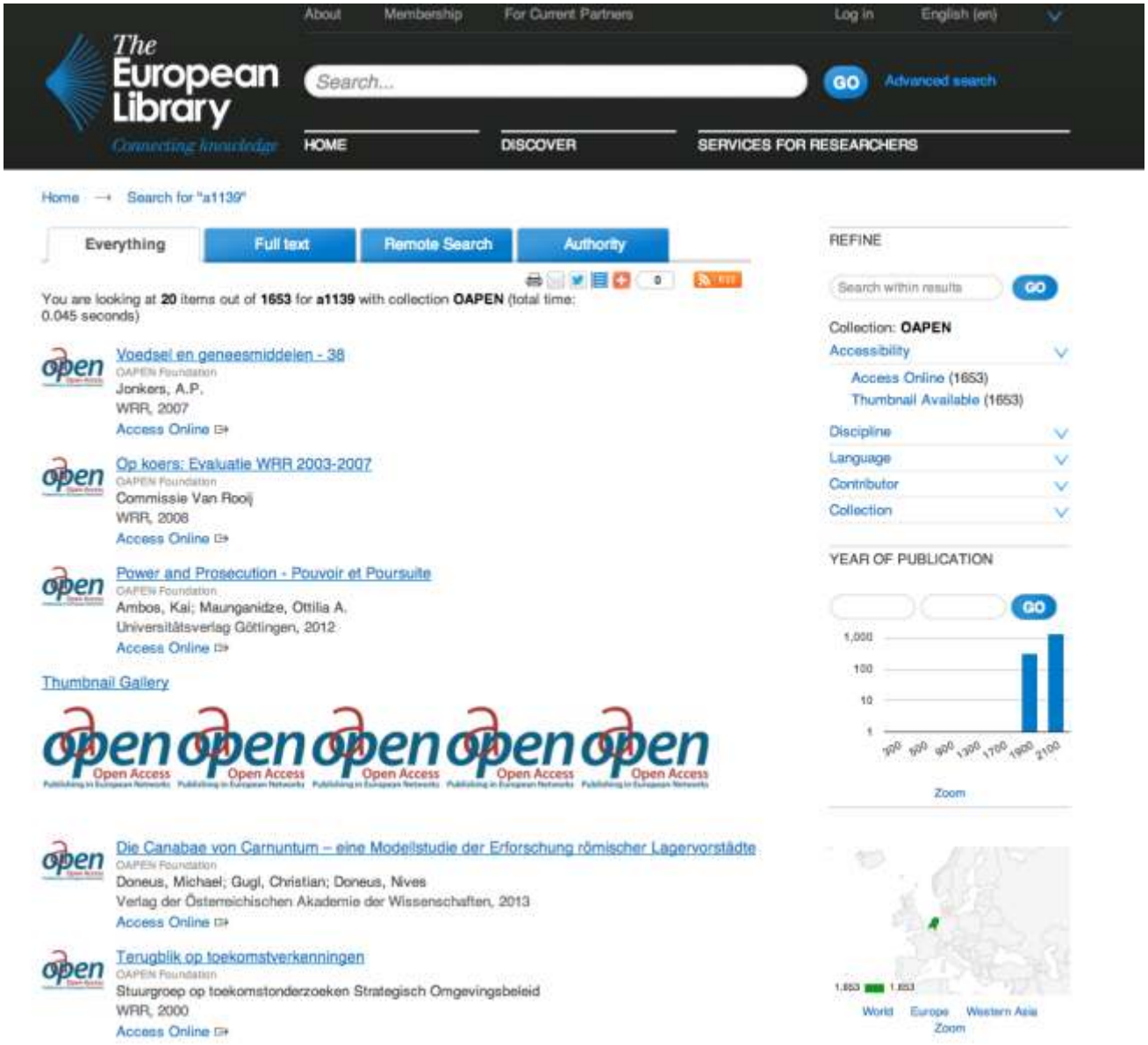
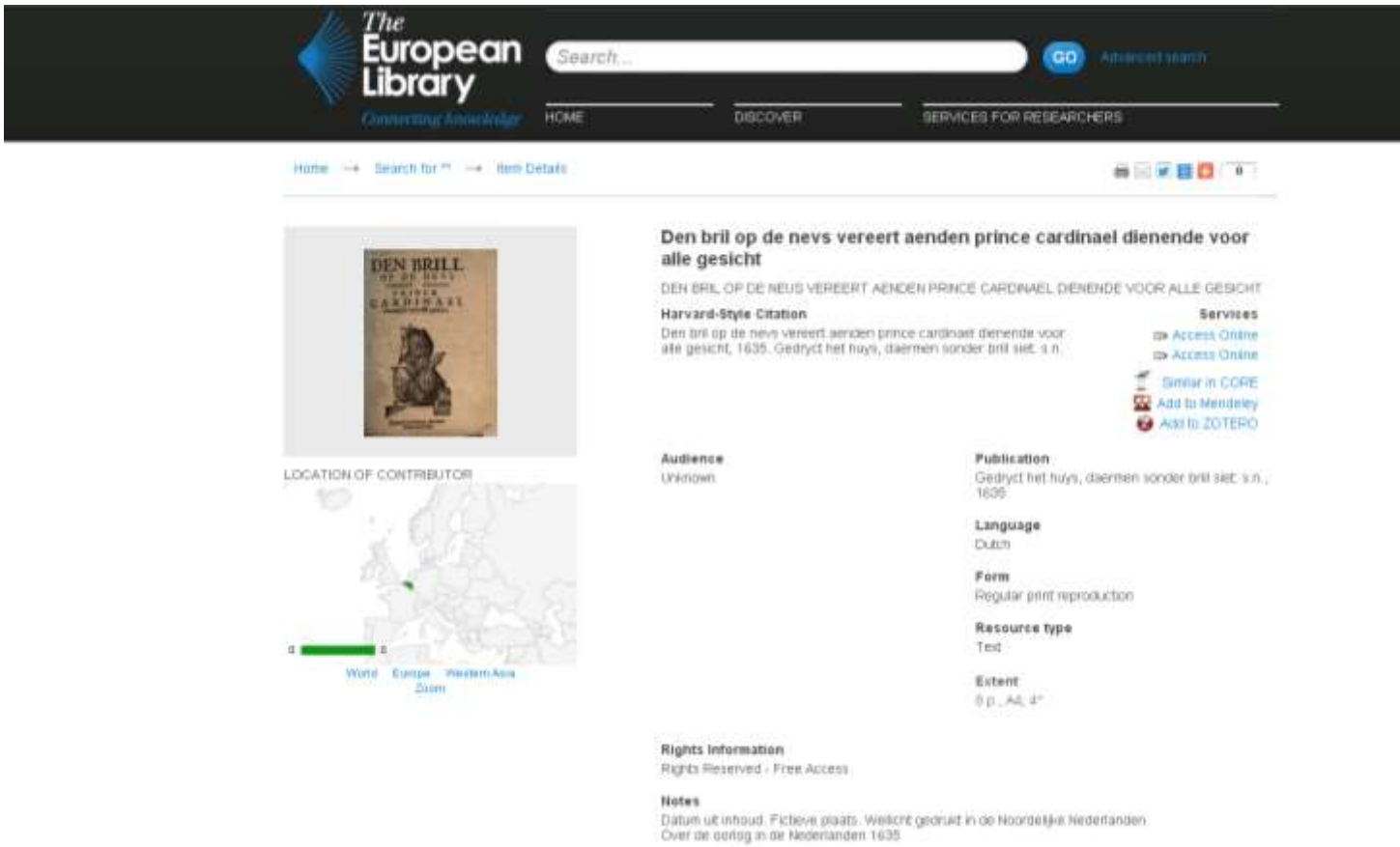
Repox – Harvesting of data via different means (http, ftp, oai-pmh, …)



Unified Ingestion Manager (UIM) – Loading, normalization, enrichment and indexing workflows to bring data from Repox to The European Library Portal


The European Library
TRAC
The European Library
Ingestion Control Panel
Monitoring
Active Ingestions
Finished Ingestions
Managing
Validation
Authority
Link/Field Checker
Active Ingestions
This view shows the current running ingestion activities together with their progress and a cancel button!
ID
Name
Start Time
Progress
Log
Ingestion Activity Details
Name
Dataset
Workflow
Progress
0%
Start Time
Scheduled
Completed
Failed
Total
Pause
Cancel

Acceptance portal (http://www.tel.ulcc.ac.uk/acceptance/)



Live portal (http://www.theeuropeanlibrary.org/***)***