



D5.1.2: Periodic Report on Content Ingestion Month 18

Authors:

Gunnar Urtegaard (NRA)

Kate Fernie (2Culture)

Runar Bergheim (AVINET)

Silvia Alfreider (NRA)

Revision: [M18 update]



LoCloud is funded by the European Commission's
ICT Policy Support Programme



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This deliverable contains original unpublished work except where clearly indicated otherwise.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	Introduction	4
2	Content action plans and timetable	5
2.1	The collection workflow	5
2.2	The partners' timetable	6
3	Event log	11
3.1	Overview of the Event log.....	11
3.2	Updating the Event log: content providers perspective	11
3.2.1	User accounts and Log in	11
3.2.2	Partner information	12
3.2.3	Information that is stored about collections	13
3.2.4	Adding events to collections	15
3.2.5	Events in the life-cycle of a collection.....	16
3.2.6	Reports from the event log.....	17
4	MINT	Error! Bookmark not defined.
4.1	MINT reports and metadata statistics	18
4.2	Partner activity on MINT.....	18
4.2.1	Data imports	18
4.2.2	Data transformations	19
4.2.3	Dataset publications	20
4.3	MINT metadata statistics	21
5	MoRE reports	23
5.1	Metadata sources	23
6	Metadata quality.....	27
7	Conclusions	32

1 Introduction

An important goal for LoCloud is to increase significantly the quantity of high quality digital content accessible through Europeana, provided by small and medium sized cultural institutions. Many small, local institutions have limited IT infrastructure and lack either the requisite staff skills in digitization and digital libraries. LoCloud is exploring the potential of cloud computing technologies for Europeana, both in the direction of an easier to use and a more efficient infrastructure and in the creation of a range of software services that benefit content providers and users.

During the first eighteen months of the project LoCloud has been putting in place the infrastructure and services to support the delivery of content to Europeana. At month eighteen the services that have been put in place include the MINT metadata ingestion tool and the MoRE repository. A series of micro-services have been developed and tested including tools to support small independent providers who use Wikimedia to publish their content online. A lightweight digital library system for small and medium sized institutions is under development and due for completion at month 21. Tools to prepare content online with embedded EDM metatags ready for capture by crawler services are also under development and due by month 21.

The first LoCloud content was delivered to Europeana for test harvesting during month eighteen and is due for publication in October 2014.

The project is putting in place tools to support content partners and launched a support portal and help desk during month 16. A series of training workshops on LoCloud services will take place during autumn 2014.

This deliverable, D5.1, focuses on monitoring the preparation and ingestion of metadata by LoCloud partners for Europeana.

LoCloud is using a modification of the Events Log system developed in Europeana Local to monitor progress.

2 Content action plans and timetable

The expected progress in terms of amount and types of content is 4 million items by end of year 3. The project plan anticipated that no content would be provided to Europeana during year 1, with harvesting starting during year 2 and continuing until the end of year 3.

During May and June 2014 the partners' content action plans and timetable for ingestion were reviewed. The results of this monitoring activity were published in an appendix to D1.4 - Consolidated Action Plans.

2.1 The collection workflow

The Partners in LoCloud are aware that it is important to follow the workflow of a collection from LoCloud provider to Europeana and provide important data from each phase.

There is no universal definition of the term collection. In LoCloud collection is used about the content listed in the original project application – as well as content from any new data sources associated with the project throughout the execution period. One database/dataset with a uniform structure can hold more than one digital collection in a more traditional use of the term. However, from a technical perspective, a digital collection is all the data which can be extracted from the same set of tables in one database, carrying the same set of attributes – without heed to thematic divisions based on content types, themes etc.

Once a LoCloud partner has reached agreement with a collection “owner” (this may be a local cultural institution within their region or a department in their organisation) to provide the content to Europeana, the following actions need to be completed for each collection:

1. The Europeana data exchange agreement (or and equivalent agreement with an aggregator) needs to be signed
2. Any IPR in the content needs to be cleared
3. Metadata needs to be captured and/or enriched (enrichment is optional)
4. Unique identifiers or URIs need to be provided for the content and metadata
5. The content needs to be published online with its metadata
6. A means of exporting metadata for aggregation needs to be established, this includes:
 - a. Choosing a system for managing the content and metadata – a digital library, repository or other system
 - b. Deciding whether to export metadata in an XML file for upload, or to establish an OAI-PMH repository, or some other means by which the metadata can be harvested
 - c. Deciding which metadata schema to specify as the output format.

LoCloud has established an Event Log (see section 3 below), and each collection being provided to Europeana through LoCloud must be registered in the Event Log. The log is then used to record activities that take place as the collection moves through the LoCloud workflow. The following activities are documented at collection level:

- When the collection is added to the content survey in the Event Log.
- When metadata is extracted about all items in a collection

- When metadata is and mapped to a registered LoCloud intermediary schema using the MINT tool. (i.e. the process of “reading” metadata from local collection management systems, mapping them to a target metadata profile and writing them into a format readable by the chosen repository technology).
- When metadata is normalized. (i.e. the process of transforming metadata from the original notation to EDM or another standard schema format accepted by LoCloud).
- When metadata is published on MINT to the LoCloud MoRE repository.
- When the normalized metadata is harvested and ingested in the LoCloud MoRE repository.
- When metadata is enriched.
- When metadata is validated as being ready for harvesting by Europeana.
- When collection metadata are harvested from the LoCloud aggregator by Europeana.
- When the collection is added to Europeana and is available to end-users.

2.2 The partners’ timetable

A timetable for the aggregation of partners’ content was prepared following the review of partners’ content action plans in May and June 2014 (see table below). Please note that the timetable, particularly for 2015, is provisional and timings may vary slightly dependent on successful testing of the tools, and of the partners content. Quality assurance may mean that harvesting is repeated to allow partners to make corrections to metadata mappings or to enrich their metadata.

Progress with this timetable is monitored in this deliverable.

Task	2013 - May 2014	Jun- 14	Jul-14	Aug- 14	Sep- 14	Oct-14	Nov- 14	Dec- 14
All Content partners: identifying and preparing content and metadata for aggregation	ongoing: timetable varies for individual collections							
Testing of the LoCloud infrastructure, microservices and Lightweight Digital Library								
UoY ADS: MINT-> MORE			MINT					
Zavad Jara: direct harvest			test		monthly harvesting			
CUT: direct harvest			test					
PSNC: direct harvest			test					
IPCHS: MINT-> MORE					test			
MECD: direct harvest + LDL						test		
BJC: direct harvest						test		
PL: direct harvest			test					
CG33: direct harvest			test					
KUAS: MINT-> MORE						MINT	test	
DP: MINT-> MORE						MINT		
PSRL: direct harvest + LDL						test		
FRS: MINT-> MORE						MINT training		

NPU: MINT-> MORE								MINT
VUKF: MINT-> MORE								MINT
AIT: direct harvest							test	
ABMR: direct harvest							test	
BGB: MINT-> MORE								test
NRA: MINT -> MORE		MINT						
AHAI: MINT-> MORE + LDL						MINT		
PrifUK KAEG: MINT-> MORE								MINT
FMNF: MINT-> MORE								
RCE: MINT-> MORE						MINT		
HU: MINT-> MORE						MINT training		
GKR: direct harvest								
Future Library: LDL								
UDE: to be confirmed								

Towards the end of the reporting period content from CUT and Zavad Jara was provided to Europeana for test harvesting, and is due to be published in October 2013.

A report is available in the Event log showing the planned harvest date for each collection that the partners will provide. This report is updated in real time as details are added by partners and is available at: http://locloudlog.avinet.no/report_collections_by_planned_ingestiondate_list.aspx

Partner	Collection	Harvest date	Digital objects
CUT	Archive of Limassol Municipality	8.2014	500
CUT	POSTAL SERVICES	8.2014	1100
PSNC	Teatr NN	9.2014	18300
FRS	Porcelain	9.2014	661
FRS	Textiles	9.2014	274
FRS	Pictures	9.2014	1000
FRS	Prints	9.2014	2800
	Local history collections, newspapers and		
GKR	magazines	9.2014	100
GKR	Local heritage Crikvenica Collection Postcards	9.2014	500
FRS	Maps	9.2014	130
FRS	Paintings	9.2014	150
FMNF	Museum collection	10.2014	
KUAS	Regin tool for 100 museum collections	11.2014	200000
CUT	Church of Cyprus	11.2014	4
CUT	Church of Cyprus	11.2014	2000
CUT	CY Police Forces	11.2014	100
CUT	Cyprus Broadcasting Corporation	11.2014	1
CUT	Ministry of Agriculture	11.2014	100
CUT	Press and Information Office CY Government	11.2014	100
	Don Juan Archive Vienna: theatre related		
AIT	texts	12.2014	4000
AIT	Hugo Montfort Digital Edition	12.2014	100

Partner	Collection	Harvest date	Digital objects
AIT	Numismatic Collection at the University of Graz	12.2014	3500
AIT	University of Graz: Archaeological Collections	12.2014	500
AIT	Visual Art of South-Eastern Europe	12.2014	3000
BGB	Belgrade local history	12.2014	1550
NPU	VAL - significant archaeological sites	12.2014	6350
ABMR	ABF Kramfors	12.2014	
ABMR	Ulvöarnas kulturarv	12.2014	
ABMR	Ånge kommun (Ånges fotosamling)	12.2014	
UoY-ADS	Grey Literature Library	12.2014	12000
UoY-ADS	Image Bank	12.2014	450
UoY-ADS	PSAS - Proc Soc Ant Scotland	12.2014	4000
UoY-ADS	Southampton Museum's Archive	12.2014	424
UoY-ADS	Star Carr Archive	12.2014	2500
UoY-ADS	Wessex Archaeology Archive	12.2014	300
	geophysical images of buried archaeological structures (buildings, chapels..), photos of castels, strongholds, fortifications, historical and cultural buildings		
PrifUK KAEG		12.2014	1000
VUKF	Society of Lithuanian Archaeology	12.2014	15000
FMNF	Trains of Portugal	1.2015	
IPCHS	Archaeological research database	4.2015	7000
Future Library	Chania	6.2015	
Future Library	Drama	6.2015	
Future Library	Ilioupoli	6.2015	
Future Library	Keratsini-Drapetsona	6.2015	
Future Library	Korinth	6.2015	
Future Library	Kozani	6.2015	
Future Library	Levadia	6.2015	
Future Library	Naupaktos	6.2015	
Future Library	Trikala	6.2015	
NPU	SOVAMM	8.2015	500
BJC	Local libraries in Cluj County		1000
BGB	New content providers		
CG33	Archives of Gironde collections		2411225
CG33	Estuary of Gironde network		5000
CG33	New collections		
Discovery	Archeaological research images		5000
Discovery	Church Body Library		
Discovery	Discovery Programme image collection		
Discovery	Dublin Institute of Advanced Studies		
Discovery	Hunt Museum Limerick		2000
Discovery	Irish Architectural Archive images		1000
Discovery	Leo Swan Aerial Collection		
Discovery	Members of AARG, private collections of archaeological sites		

Partner	Collection	Harvest date	Digital objects
Discovery	Royal Society of Antiquaries of Ireland		15000
FMNF	Archives collection		
FMNF	Municipality of Abrantes		
FMNF	Municipality of Vila Nova de Famalicão		
FMNF	REFER - Heritage Department		
GKR	Local museum items		500
HU	Vehbi Koç & Ankara Research Centre (VEKAN)		1000
Zavod Jara	KAMRA		6164
Zavod Jara	User generated content		400
IPCHS	Works of art database, restoration		
RCE	Archaeological depot, province of Gelderland		
RCE	Archaeological reports		
RCE	Controlled vocab of Dutch archaeology (RNA)		
RCE	Historic Cultural Landscapes		
	Other archaeological depots and museums		
RCE	(not present in DiMCoN)		
RCE	Regional/local museums		
MECD	Biblioteca Central de Cantabria		
MECD	Biblioteca Miguel Hernández - La Zubia		
MECD	Biblioteca Municipal de Caravaca de la Cruz		
MECD	Biblioteca Municipal de Catral		
MECD	Biblioteca Municipal de Ermua		
MECD	Biblioteca y Archivo Municipal de Monóvar		
MECD	Bibliotecas Públicas de Mairena de Aljarafe		
MECD	Fundación Juanelo Turriano		
MECD	Institute for Spanish Cultural Heritage		15000
	Museo Arqueológico de Linares. Monográfico		
	de Cástulo		
MECD	Museo Arqueológico Provincial de Ourense		
MECD	Museo Casa de los Tiros		
MECD	Museo de Albacete		
	Museo de Artes y Costumbres Populares de		
MECD	Sevilla		
	Museo de Artes y Costumbres Populares del		
MECD	Alto Guadalquivir		
MECD	Museo de Cádiz		
MECD	Museo de Ciudad Real		
MECD	Museo de Cuenca		
MECD	Museo de Guadalajara		
MECD	Museo de Historia		
MECD	Museo de Huelva		
MECD	Museo de Huesca		
MECD	Museo de Jaén		
MECD	Museo de las Peregrinaciones y de Santiago		
MECD	Museo de los Orígenes		
MECD	Museo de Málaga		

Partner	Collection	Harvest date	Digital objects
MECD	Museo de Mallorca		
MECD	Museo de Santa Cruz		
MECD	Museo de Teruel		
MECD	Museo de Zaragoza		
	Museo del Parque Cultural de Molinos.		
MECD	Colección Eleuterio Blasco Ferrer		
MECD	Museo Etnológico de Ribadavia		
MECD	Museo Juan Cabré		
MECD	Museo Martín Almagro de Albarracín		
MECD	SG for library coordination		95000
MECD	SG for state museums		20000
MECD	Shadous: memoria fotográfica		
AHAI	Archaeological sites		1200
AHAI	Municipality of Skagafjorour, local museum		
NRA	Local history		2020000
PSRL	Maps		100
PSRL	Postcards		800
PSRL	Seals		100
Provincie Limburg	Placenames		10000
Provincie Limburg	Various		50000
ABMR	Birgittamuseet - medicine history		
ABMR	Kubikenborgs skolas intresseförening		
ABMR	Landsarkivet		
UDE	Center for International Light art in Unna		
UDE	City Archives of Witten		
UDE	Conserve the Sound Essen		
UDE	Gustav-Lübcke-Museum Hamm		
	Kunstmuseum Mülheim an der Ruhr in der		
UDE	Alten Post		
	Local association Hattingen/Ruhr / Museum		
UDE	in the Iron House		
	Museum of the city Gladbeck in the Water		
UDE	tower Wittringen		
	The Ruhr-University Bochum - Modern Art		
UDE	Gallery		
PrifUK KAEG			
PrifUK KAEG			
VUKF	Archaeological thesaurus		
VUKF	Historic placenames thesaurus		
Total digital objects			4950490

3 Event log

LoCloud has put in place a modification of the Events Log system used in EuropeanLocal; the system has been set up by Avinet. Data about partners' collections captured during the content survey carried out in April-May 2014 has been ingested into the system, which was launched as a fully functional service in September 2014, in time for monitoring at month 18.

In the context of this tool, the term event relates to when an activity in the LoCloud workflow takes place on an individual collection and is intended to allow progress through the workflow to be monitored.

The event log is online at: <http://locloudlog.avinet.no>

3.1 Overview of the Event log

The Event Log is a simple database with online data entry forms, reporting and analyzing functions.

When a content provider or aggregator has done work on one or several collections, he/she will connect to the Event Log, select the collection involved and add information from the predefined list of events. At a minimum this can be just some very simple core facts that will take a few minutes to add, sufficient to document the action and enter amount of items and objects and date the work was done. This will be enough to maintain data about amount and progress.

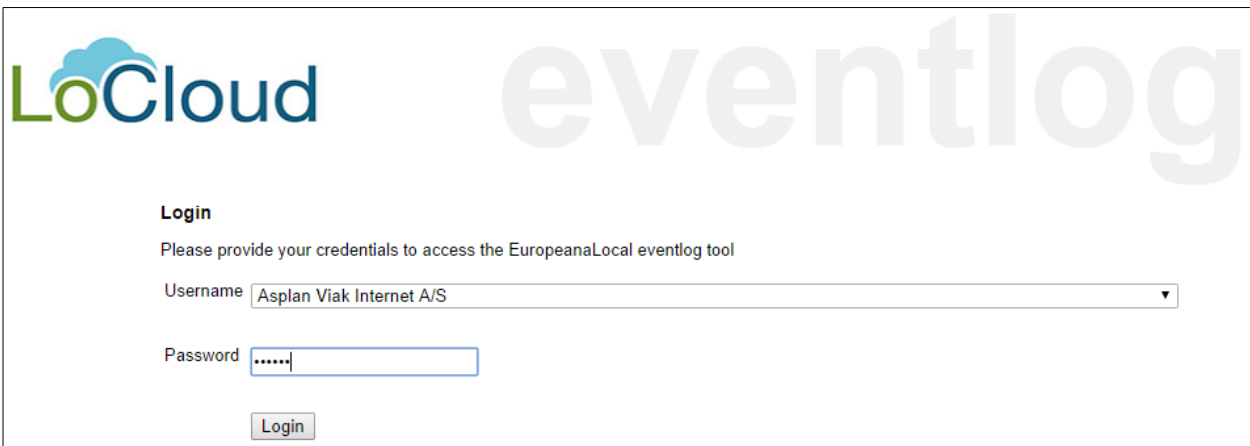
If the content provider or the aggregator encounters any problems or wishes to share lessons learned and ideas about improving the processes, this can be added as a manual report and thus being made available to the rest of the partners in the project.

3.2 Updating the Event log: content providers perspective

3.2.1 User accounts and Log in

User accounts have been set up for each partner to enable them to report on progress as the collection moves through the workflow. These accounts are shared with all of the LoCloud support services.

The log-in screen provides a drop-down list of the organisation names of registered users to simplify the process.



LoCloud eventlog

Login
Please provide your credentials to access the EuropeanaLocal eventlog tool

Username

Password

3.2.2 Partner information

Once partners are logged into the system they will see a list of options, which allow them to view their collections, a series of reports and their partner details.

The “Add & Edit” section includes the option to select “Partner”.

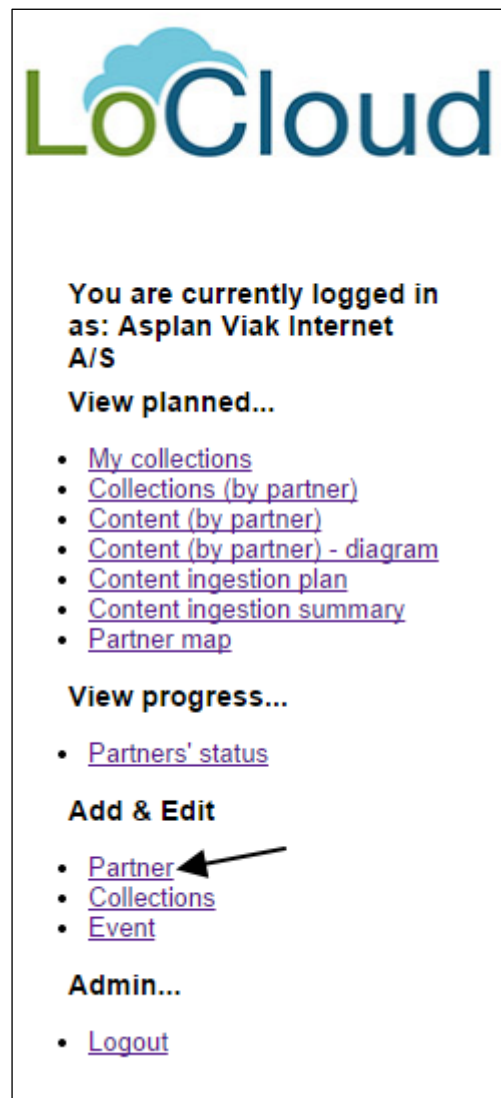
This displays a list of all partners that the user is authorized to edit. This may be limited to their own organization, but can include other organisations in their region. The administrator view (figure below) shows all registered organisations.

Add or edit partner information

On this page you can add new partners or edit information about regist

[Add new](#)

Select	Country	Partn
	AT	Angewandte Informationstechnik Forschungsgesellsch
	BE	Provincie Limburg
	BG	Pencho Slaveykov Regional Library
	CY	Cyprus University of Technology
	CZ	Národní památkový ústav (National Heritage Institute)
	DE	Universitaet Duisburg-Essen (University of Duisburg-E
	DK	Kulturarvsstyrelsen (Danish Agency for Culture)
	ES	Ministerio de Educacion, Cultura y Deporte (Ministry o
	FR	Conseil Général de la Gironde - Archives Départemen
	GR	Future Library
	HR	Gradska knjižnica Rijeka
	IE	Discovery Programme
	IS	Minjastofnun Íslands /The Cultural Heritage Agency of
	IT	Fondazione Ranieri di Sorbello



LoCloud

You are currently logged in as: Asplan Viak Internet A/S

View planned...

- [My collections](#)
- [Collections \(by partner\)](#)
- [Content \(by partner\)](#)
- [Content \(by partner\) - diagram](#)
- [Content ingestion plan](#)
- [Content ingestion summary](#)
- [Partner map](#)

View progress...

- [Partners' status](#)

Add & Edit

- [Partner](#) ←
- [Collections](#)
- [Event](#)

Admin...

- [Logout](#)

Updating the details is a simple process of selecting the partner to update by clicking on the green button in the first column of the table. The details about the organization are then displayed at the top of the page. Clicking the “edit” link allows the form to become editable.

Country	Norway
Region	N/A
Partner name	Asplan Viak Internet A/S
Password	*****
Plan to use MINT	<input checked="" type="checkbox"/>
Plan to use MORE	<input checked="" type="checkbox"/>
Plan to use LoCloud Collections (LDL)	<input type="checkbox"/>
Plan to test microservices	geo,enrich
Edit	
Add new	

The details include the organisation name, country and region and an indication of which of the LoCloud services the partner plans to use.

Users can update their username and also their password - but we do not encourage this, as we would like partners to maintain the same login across all LoCloud systems.

The longitude and Latitude of each partner's home office is stored to enable a "partner map" to be created.

Country	Norway
Region	N/A
Partner	Asplan Viak Internet A/S
Password	*****
Plan to use MINT	<input checked="" type="checkbox"/>
Plan to use MORE	<input checked="" type="checkbox"/>
Plan to use LoCloud Collections (LDL)	<input type="checkbox"/>
Plan to test microservices	geo,enrich
Longitude	
Latitude	
Update	
Cancel	
Delete	
Add new	

3.2.3 Information that is stored about collections

Clicking on "Collection" under the "Add & Edit" heading displays a list of all collections belonging to the user.

Add or Edit Collections

This form allows you to add new collections or edit collections belonging to the currently logged in partner.

Selected collection details

Please select a collection below to edit it or [click here to add a new collection](#)

Collections

Select	Collection name
<input type="radio"/>	LoCloud Test collection


Each collection can be selected by clicking on the green button in the first column of the table. This displays the collection information at the top of the page.

Selected collection details

Content provider	AVINET
Collection name	LoCloud Test collection
Planned harvest date	20.09.2014
Europeana DEA signed	completed
Metadata export format	EDM
Europeana rights statement	1 The Public Domain Mark (PDM)
Thumbnails available	Yes, for all digital objects
Example URL (isShownBy/At)	http://www.avinet.no
Long term consistency of URLs	Yes, for all URLs
Description	This is a test collection for LoCloud
Number of digital objects	1 text objects 1 image objects 1 sound objects 1 video objects 1 3D objects
Edit Delete New	
Add new collection	

Clicking the “edit” link below the detailed information about the collection allows the form to become editable and the details to be updated.

Selected collection details

Content provider	<input type="text" value="AVINET"/>
Collection name	<input type="text" value="LoCloud Test collection"/>
Planned harvest date	20.09.2014 <input type="text" value=""/> 
Europeana DEA signed	completed <input type="text" value=""/>
Metadata export format	<input type="text" value="EDM"/>
Europeana rights statement	1 The Public Domain Mark (PDM) <input type="text" value=""/>
Thumbnails available	Yes, for all digital objects <input type="text" value=""/>
Example URL (isShownBy/At)	<input type="text" value="http://www.avinet.no"/>
Long term consistency of URLs	Yes, for all URLs <input type="text" value=""/>
Description	<input type="text" value="This is a test collection for LoCloud"/>
Number of digital objects	<input type="text" value="1"/> text objects <input type="text" value="1"/> image objects <input type="text" value="1"/> sound objects <input type="text" value="1"/> video objects <input type="text" value="1"/> 3D objects
<input type="text" value="Update"/> <input type="text" value="Cancel"/>	
Add new collection	

The information stored about each collection is designed to complete the “Content Contribution Form” requested by Europeana prior to ingestion. It includes:

- The name of the content provider – in some cases this is different from the organization participating in LoCloud.
- Collection name – for example, the name of the collection given in the content survey.

- Planned harvest date – the date when the partner aims to have the collection ready for harvesting by Europeana
- Whether the Europeana Data Exchange Agreement (DEA) has been signed for the collection
- What format the exported metadata from your collection is available in, i.e. EDM, ESE or similar
- The Europeana Rights Statement - a declaration of the usage rights for the content of the collection. Contributors select between 13 predefined licenses specified by Europeana: <http://pro.europeana.eu/available-rights-statements>
- Whether thumbnails are available for the objects in the collection when published in Europeana
- An example URL – an example of the URL to be provided in the EDM isShownBy or isShownAt element (i.e. a direct link to the object or a landing page for the object)
- Description – for notes about the collection or simply a description of its content.
- The number of digital objects categorized by texts, images, sounds, videos and 3d objects

3.2.4 Adding events to collections

Clicking on “Event” under the “Add & Edit” allows events to be added to each collection. The page opens displaying one of the collections belonging to the currently logged in user. A dropdown list allows users to choose which collection to work on.

Select the collection you wish to add events for

LoCloud Test collection ▾ ← Select collection here

Selected collection details

Content provider	AVINET
Collection name	LoCloud Test collection
Planned harvest date	20.09.2014
Europeana DEA signed	in progress
Metadata export format	EDM
Europeana rights statement	1 The Public Domain Mark (PDM)
Thumbnails available	Yes, for all digital objects
Example URL (isShownBy/At)	http://www.avinet.no
Long-term consistency of URLs	Yes, for all URLs
Description	This is a test collection for LoCloud
Number of digital objects	5 text objects 1 image objects 1 sound objects 1 video objects 1 3D objects

← Collection details here

Events

← List of events here

[Add new event](#)

Date	Event	Status	Information about the event	Delete
09.09.2014	IPR cleared	completed		Delete
09.09.2014	Metadata captured	completed		Delete

The page displays details about the collection and, at the bottom of the page, a list of events for the collection.

Existing events can be deleted by choosing the “Delete” link.

Events

[Add new event](#)

Date	Event	Status	Information about the event	Delete
09.09.2014	IPR cleared	completed		Delete
09.09.2014	Metadata captured	completed		Delete

New event can be added to the collection by clicking the “Add new events” link below the header “Events”. This displays the editing form for a new event.

Events

[Add new event](#)

Input date of event	09.09.2014
Select an event type	IPR cleared
Select status for the event	in progress
Comments	
Attachments	Choose File No file chosen
Insert Cancel	

Date	Event	Status	Information about the event	Delete
09.09.2014	IPR cleared	completed		Delete
09.09.2014	Metadata captured	completed		Delete

By default, the form will be filled in with today’s date – this can be changed using the date picker or by entering a date in the format dd.mm.yyyy.

3.2.5 Events in the life-cycle of a collection

There are eight pre-defined events in the life-cycle of a collection; these are:

1. IPR cleared
2. Metadata captured
3. Metadata enriched
4. Content published
5. Persistent URIs established
6. Metadata exported
7. Metadata harvested
8. Validation completed

For each collection, we would like you to fill in the information as the various stages are reached. Each event has a date and a status that can be either “in progress” or “completed”. Any other value will be assumed to be not done or incomplete.

Users can write a comment related to the event and upload an attachment if relevant, but this is not mandatory.

3.2.6 Reports from the event log

The Event Log is an important tool for collecting and sharing information about the amount of content and progress in LoCloud. Reports can be selected by country, region, type of content, time span, amount etc.

The Event Log documents what happened to each collection. It will enable many interesting analyses of the data to be performed, for example:

- The number of items and objects harvested from local providers at any given time. A report organized by country will be updated on a daily basis by the system.
- The number of collections, items and objects are entered in the Event Log at any given time? (arranged by country, domain etc).
- How many digital objects from a chosen number of collections are extracted and mapped by a given date.
- How many items were affected by metadata normalization and enrichment from a country, a region, certain types of collections etc in a given period of time.
- How many items and objects are harvested into LoCloud repository by any date, country, region, provider type (museum, archive, library).
- Items and objects harvested by Europeana by end of year 2, end of year 3.
- The number of providers and collections involved in LoCloud at any given time.
- The number of manual reports added, from which country, provider, etc.

4 MINT

The MINT mapping tool (<http://mint-projects.image.ntua.gr/locloud/>) is a web based platform designed to support metadata mapping and ingestion for cultural heritage content and metadata in Europe. The main activities by users of MINT are:

- Registration
- Metadata import
- Creating a metadata mapping
- Metadata transformation
- Publication of data to the LoCloud MoRE repository

4.1 MINT reports and metadata statistics

MINT includes reporting functionality that provides information about users' actions and their progress in MINT. These reports allow the following actions to be monitored:

1. Dataset Imports.
2. Dataset Transformations.
3. Dataset Publications.
4. An overview of each Organization's progress by dataset with item counts (imported, transformed, published).
5. A Project overview of all the datasets imported, transformed or published by each partner organization.

4.2 Partner activity on MINT

By the end of August 2014, excluding NTUA, 7 partners had imported datasets to MINT. The statistics suggest that this was to test the mapping tool, or to analyse their data output.

4.2.1 Data imports

Partner	Name	Date	Items	Number of imports	Total imported items
ABMR	http://repor.murberget.se:	19/03/2014	34143	1	34143
ADS	amarna_for_import-corrected2.xml	25/07/2014	40	1	40
NRA	MUSIT photos Oslo 1000 rec.xml	02/07/2014	1000	3	6782
NRA	Balestrand_foto2.csv	28/06/2014	4782		
NRA	MUSIT artefacts Oslo 1000 rec.xml	24/06/2014	1000		
NPU	http://iispp.npu.	27/08/2014	0	2	117

Partner	Name	Date	Items	Number of imports	Total imported items
BJC	http://greenstone.bjc.	03/06/2014	10	1	10
FRS	output2_xml.zip	22/07/2014	1470	15	4422
GKR	http://svevid2.codingthrills.	11/06/2014	73	1	73

The name of the data import shows that 4 partners (ABMR, NPU, BJC and GKR) provided links to their OAI-PMH repositories, which were then harvested by MINT. One partner (FRS) uploaded metadata in zip files and 3 partners uploaded metadata as XML files, with 1 also uploading data in a CSV file.

4.2.2 Data transformations

Data transformation is the process that takes place once a mapping has been completed between the metadata structure of the imported dataset and a target schema. The report on data transformations shows both the target schema and mapping used, and also the number of items that, once transformed, conform to the schema (are valid according to the schema's rules).

The report shows that the 3 partners that have transformed their datasets have completed 15 different data mappings, in 2 cases experimenting with mappings to different target schemas. There is also some suggestion from the statistics (the number of invalid items and repeated mappings and transformations) which that the partners involved were testing the validity of their datasets – perhaps in order to make adjustments to their data output, at the same time as testing the process.

P	Name	Items	Valid	Invalid	Target schema	Mapping used	Parent Import Name
NPU	CARARE 2.0.1 Transformation	117	117	0	CARARE 2.0.1	Pamatky2-	http://iispp.npu.
NRA	CARARE 2.0.3 Transformation	1000	743	257	CARARE 2.0.3	MUSIT photos test	MUSIT photos Oslo 1000
NRA	EDM Transformation	4782	0	4782	EDM	joachim_test	Balestrand_foto2.csv
NRA	CARARE 2.0.3 Transformation	1000	323	677	CARARE 2.0.3	MUSIT artefacts test	MUSIT artefacts Oslo 1000
FRS	EDM Transformation	1470	1080	390	EDM	items	output2_xml.zip
FRS	EDM Transformation	1470	1470	0	EDM	output2bis	output_xml.zip
FRS	EDM Transformation	1470	1080	390	EDM	items	output_xml.zip
FRS	EDM Transformation	1	0	1	EDM	78388	oaiexport_oa.xml78388.
FRS	EDM Transformation	1	0	1	EDM	81654	oaiexport_oa.xml78388.

P	Name	Items	Valid	Invalid	Target schema	Mapping used	Parent Import Name
FRS	EDM Transformation	1	1	0	EDM	81654	oaiexport_oa.xml78380.
FRS	EDM Transformation	1	0	1	EDM	81654	oaiexport_oa.xml78379.
FRS	LIDO-Locloud Transformation	1	0	1	LIDO-Locloud	81656	oaiexport_oa.xml78379.
FRS	LIDO Transformation	1	0	1	LIDO	78378lido	oaiexport_oa.xml78378.
FRS	EDM Transformation	1	0	1	EDM	78378bis	oaiexport_oa.xml78378.
FRS	EDM Transformation	1	0	1	EDM	78378bis	oaiexport_oa.xml78378.
FRS	EDM Transformation	1	0	1	EDM	78378	oaiexport_oa.xml78378.
FRS	EDM Transformation	1	0	1	EDM	78378	oaiexport_oa.xml78378.
FRS	EDM Transformation	1	0	1	EDM	78378bis	oaiexport_oa.xml78377.
FRS	EDM Transformation	1	1	0	EDM	78376	oaiexport_oa.xml78376.
FRS	EDM Transformation	1	1	0	EDM	81654	oaiexport_oa.xml81657.
FRS	EDM Transformation	1	1	0	EDM	81654	oaiexport_oa.xml81654.
FRS	EDM Transformation	1	0	1	EDM	oa.xml81657	oaiexport_oa.xml81657.
FRS	prova xsl Transformation	0	0	0		prova	oaiexport_oa.xml78377.

4.2.3 Dataset publications

Once partners have completed their activities on MINT and are satisfied that their dataset transformations are valid, the final step is publication. Publication is the process by which datasets are packaged with their mappings and sent to the LoCloud MoRE repository for ingestion.

There were no datasets published during the period.

4.3 MINT metadata statistics

In addition to reporting on user’s activities, MINT also provides tools to analyse the statistics for individual datasets. The figure below illustrates the statistics for the dataset uploaded by ABMR, showing the element names, the count for each element, the number of distinct values for each element and the average length of the data contained in each element.

<http://repop.murberget.se:8080/repop/OAIHandler> - KAP_Handel_lido Statistics

Xpath	Count	Distinct	Length
lido:lidoWrap	34143		
@xsi:schemaLocation	34143	1	79
lido:lido	34143		
@xsi:schemaLoca...	34143	1	79
lido:lidoRecID	34143	34143	18.7
@lido:source	34143	1	24
@lido:type	34143	1	5
lido:objectPublishedID	34143	34143	55.7
@lido:type	34143	1	3
lido:category	34143		
lido:conceptID	34143	1	42
@lido:type	34143	1	3
lido:term	34143	1	15
+xml:lang	34143	1	2
lido:descriptiveMetad...	34143		
+xml:lang	34143	1	2
lido:objectClassifi...	34143		
lido:objectWork...	34143		
lido:objectW...	34143		
lido:term	68286	2	8
@lid...	68286	1	3
@xm...	68286	2	2
lido:classificati...	34143		
lido:classific...	102429		
@lido:t...	102429	3	12.3
lido:term	102429	3	10.3

By looking at the dataset statistics, it is possible to see if core information is present in the dataset. The core data required for Europeana is:

- Title or Description
- Subject, Type, Coverage or Spatial
- Rights licencing
- A URL or URI pointing either directly to the object or to a landing page where it can be found

This requirement means that to be valid for Europeana, each item in the dataset must contain relevant data in at least 4 elements.

The statistics provided in MINT allow for some additional checks to be made. For example, each record needs a unique identifier – an element which has the same number of distinct values as in the total count for the dataset can provide a unique identifier.

The tool also allows the data that is actually stored in an element to be browsed, which can be very helpful when mapping a dataset but can also provide a way of monitoring the quality. For example users can review the consistency and usefulness of the data actually present.

▼ **lido:displayActor information**

XPath /lidoWrap/lido/descriptiveMetadata/objectRelationWrap/subjectWrap/subjectSet/subject/subjectActor/displayActor

Namespace URI http://www.lido-schema.org

Count 34757

Distinct Count 30021

Average Length 18.102425

Value	Count
Centralhotellet	14
Byggnadsfirman	12
Fruktkällaren	12
Ingenjörfirma	12
Nya Damfriseringen	12
Alltjänst	11
J. Näslund	11
J.A. Johansson	11
Nya Modemagasinet	11
Byggnadsfirma	10
Centralaffären	10
E. Eriksson	10
Erik Johansson	10
J. Andersson	10
Nya Modeaffären	10

Prev 1 2 3 4 5 ... 2002 Next

In this example, the element called “actor information” appears to contain a mixture of people’s names and place/organisation names.

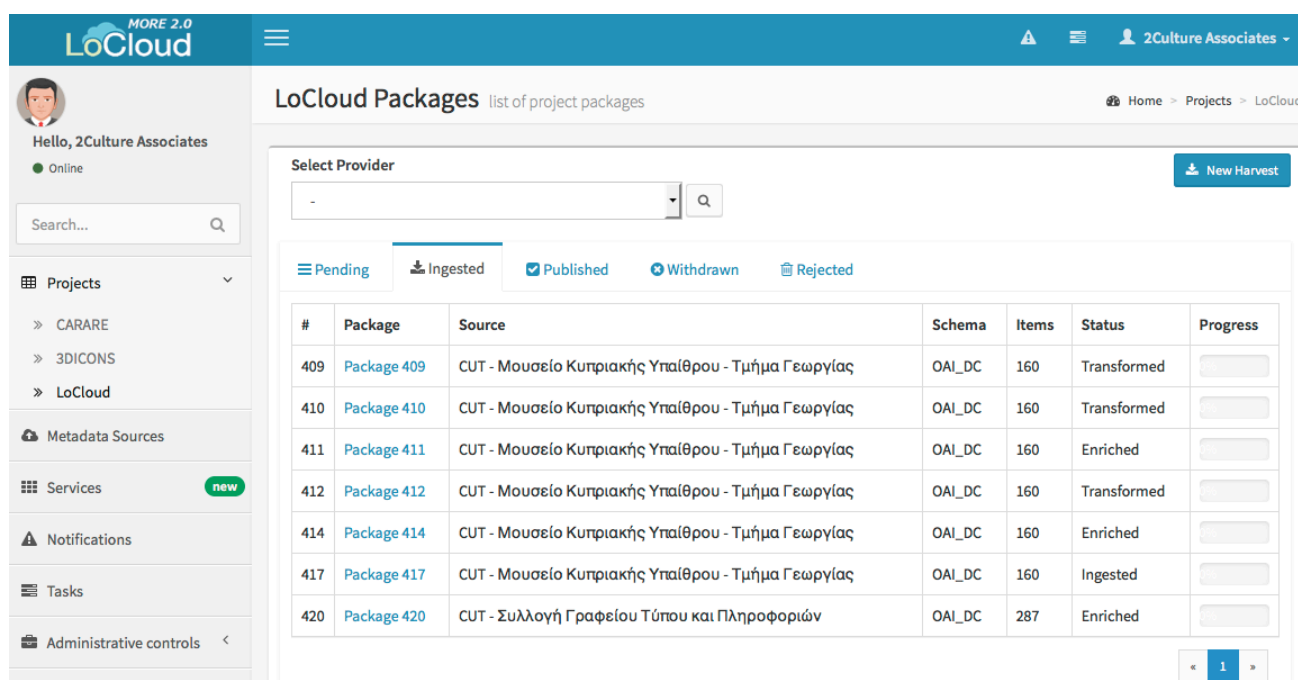
5 MoRE repository

The MoRE repository (<http://store.locloud.eu>) is a web-based platform to support metadata harvesting, ingestion, validation, enrichment and delivery to Europeana. The main activities on MoRE are:

- Harvesting metadata sources including MINT and repositories maintained by LoCloud partners
- Metadata ingestion
- Metadata enrichment
- Metadata transformation to EDM
- Publication of data to Europeana.

5.1 Metadata ingested

The MoRE interface allows administrators to see which packages of content are pending ingestion and which have been ingested.



The screenshot shows the LoCloud Packages interface. The header includes the LoCloud logo, a user profile for '2Culture Associates', and navigation links. The main content area displays a table of packages with columns for ID, Package name, Source, Schema, Items, Status, and Progress. The 'Ingested' tab is selected, showing a list of packages with their respective details.

#	Package	Source	Schema	Items	Status	Progress
409	Package 409	CUT - Μουσείο Κυπριακής Υπαίθρου - Τμήμα Γεωργίας	OAI_DC	160	Transformed	<div style="width: 100%;"></div>
410	Package 410	CUT - Μουσείο Κυπριακής Υπαίθρου - Τμήμα Γεωργίας	OAI_DC	160	Transformed	<div style="width: 100%;"></div>
411	Package 411	CUT - Μουσείο Κυπριακής Υπαίθρου - Τμήμα Γεωργίας	OAI_DC	160	Enriched	<div style="width: 100%;"></div>
412	Package 412	CUT - Μουσείο Κυπριακής Υπαίθρου - Τμήμα Γεωργίας	OAI_DC	160	Transformed	<div style="width: 100%;"></div>
414	Package 414	CUT - Μουσείο Κυπριακής Υπαίθρου - Τμήμα Γεωργίας	OAI_DC	160	Enriched	<div style="width: 100%;"></div>
417	Package 417	CUT - Μουσείο Κυπριακής Υπαίθρου - Τμήμα Γεωργίας	OAI_DC	160	Ingested	<div style="width: 100%;"></div>
420	Package 420	CUT - Συλλογή Γραφείου Τύπου και Πληροφοριών	OAI_DC	287	Enriched	<div style="width: 100%;"></div>

Each package consists of a dataset that has been harvested from a source together with information about that dataset. The figure above shows a series of packages that have been harvested and ingested from CUT's Omeka repository in a Dublin Core metadata format, enriched and then transformed to EDM

5.1.1 Package information

More detailed information is available about each package. The figure below shows the details including the number of items in the package, the number of items that have been validated and the number transformed to EDM.

Package Information Project LoCloud Home > Projects > LoCloud > 409

[Back](#) **Package 409**

Total items: 160

Status: Transformed

Schema: OAI_DC **Source:** CUT - Μουσείο Κυπριακής Υπαιθρου

[History](#)
[Notifications](#)
[Reject Package](#)

[View Items](#)
[Enrich](#)
[Transform](#)

[Publish](#)

Statistics

Package Details

Type	Results
Harvested items	160
Validated items (structural)	160
Validated items (schema)	160
Items with validation errors	0
Transformed items	160

Completeness

Further statistics are in development for the MoRE repository.

5.1.2 View items

From the package information it is possible to view all of the items contained within.

View Items list of items for selected packages Home > LoCloud > Package 420

[Back](#) **Project LoCloud :: Package 420**

#	Item	Native ID	
1	Γεωργός στο Δάλι	oai:https://apsida.cut.ac.cy:11861	OAI_DC EDM eEDM
2	Καλλιέργειες (ψέκασμα) στην Αθηνών	oai:https://apsida.cut.ac.cy:11862	OAI_DC EDM eEDM

For each item it is possible to view the metadata in both its native format, in this case OAI_DC, and transformed to EDM. This allows for checking of the semantic content of a sample of records within the dataset.

```
- <oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:title>Γεωργός στο Δάλι</dc:title>
  - <dc:creator>
    Γραφείο Τύπου και Πληροφοριών (Κυπριακή Δημοκρατία)
  </dc:creator>
  <dc:subject>Αγροτική ζωή--Κύπρος</dc:subject>
  <dc:subject>Agriculture life--Cyprus</dc:subject>
  <dc:subject>Αγρότες--Κύπρος--Ιστορία </dc:subject>
  <dc:description>Ο γεωργός μαζεύει το σιτάρι στο χωριό Δάλι.</dc:description>
  <dc:publisher>Βιβλιοθήκη Τεχνολογικού Πανεπιστημίου Κύπρου</dc:publisher>
  - <dc:publisher>
    Εργαστήριο Ψηφιακής Πολιτιστικής Κληρονομιάς Τεχνολογικού Πανεπιστημίου Κύπρου
  </dc:publisher>
  - <dc:contributor>
    Γραφείο Τύπου και Πληροφοριών (Κυπριακή Δημοκρατία)
  </dc:contributor>
  <dc:date>1950-1960</dc:date>
  <dc:type>Image</dc:type>
  <dc:format>JPG</dc:format>
  <dc:identifier>1A-007-003</dc:identifier>
  <dc:identifier>https://apsida.cut.ac.cy/items/show/11861</dc:identifier>
  - <dc:identifier>
    https://apsida.cut.ac.cy/files/original/5e97b03af8ec01c8332f766eb7b9c8f9.jpg
  </dc:identifier>
  - <dc:source>
    Γραφείο Τύπου και Πληροφοριών (Κυπριακή Δημοκρατία)
  </dc:source>
  <dc:language>EL, EN</dc:language>
  <dc:coverage>35.021093, 33.419368</dc:coverage>
  - <dc:rights>
    Απαγορεύεται η δημοσίευση ή αναπαραγωγή, ηλεκτρονική ή άλλη χωρίς τη γραπτή συγκατάθεση του δημιουργού.
  </dc:rights>
</oai_dc:dc>
```

The figure above shows the native metadata and the figure below the metadata transformed to EDM.

```

- <edm:ProvidedCHO rdf:about="http://more.loccloud.eu/object/CUT/11861">
  <dc:title xml:lang="">Γεωργός στο Δάλι</dc:title>
  <dc:description xml:lang="">Ο γεωργός μαζεύει το σιτάρι στο χωριό Δάλι.</dc:description>
  <dc:subject>Αγροτική ζωή--Κύπρος</dc:subject>
  <dc:subject>Agriculture life--Cyprus</dc:subject>
  <dc:subject>Αγρότες--Κύπρος--Ιστορία </dc:subject>
  <dc:terms:spatial rdf:resource="http://more.loccloud.eu/CUT/11861/SP.2"/>
  <edm:type>IMAGE</edm:type>
- <dc:creator xml:lang="">
  Γραφείο Τύπου και Πληροφοριών (Κυπριακή Δημοκρατία)
  </dc:creator>
- <dc:contributor xml:lang="">
  Γραφείο Τύπου και Πληροφοριών (Κυπριακή Δημοκρατία)
  </dc:contributor>
  <dc:date>1950</dc:date>
  <dc:date>1960</dc:date>
  <dc:publisher xml:lang="">Βιβλιοθήκη Τεχνολογικού Πανεπιστημίου Κύπρου</dc:publisher>
- <dc:publisher xml:lang="">
  Εργαστήριο Ψηφιακής Πολιτιστικής Κληρονομιάς Τεχνολογικού Πανεπιστημίου Κύπρου
  </dc:publisher>
- <dc:source>
  Γραφείο Τύπου και Πληροφοριών (Κυπριακή Δημοκρατία)
  </dc:source>
</edm:ProvidedCHO>
- <edm:WebResource rdf:about="https://apsida.cut.ac.cy/files/original/5e97b03af8ec01c8332f766eb7b9c8f9.jpg">
  <dc:format>JPG</dc:format>
  - <dc:rights xml:lang="">
    Απαγορεύεται η δημοσίευση ή αναπαραγωγή, ηλεκτρονική ή άλλη χωρίς τη γραπτή συγκατάθεση του δημιουργού.
    </dc:rights>
    <edm:rights rdf:resource="http://www.europeana.eu/rights/rr-f"/>
  </edm:WebResource>
- <edm:Place rdf:about="http://more.loccloud.eu/CUT/11861/SP.2">
  <skos:prefLabel>35.021093, 33.419368</skos:prefLabel>
  </edm:Place>
- <ore:aggregation rdf:about="http://more.loccloud.eu/object/CUT/11861#aggregation">
  <edm:aggregatedCHO rdf:resource="http://more.loccloud.eu/object/CUT/11861"/>
  <edm:dataProvider>Βιβλιοθήκη Τεχνολογικού Πανεπιστημίου Κύπρου</edm:dataProvider>
  <edm:provider>LoCloud</edm:provider>
  <edm:rights rdf:resource="http://www.europeana.eu/rights/rr-f"/>
  <edm:isShownAt rdf:resource="https://apsida.cut.ac.cy/items/show/11861"/>
  <edm:isShownBy rdf:resource="https://apsida.cut.ac.cy/files/original/5e97b03af8ec01c8332f766eb7b9c8f9.jpg"/>
  <edm:object rdf:resource="https://apsida.cut.ac.cy/files/original/5e97b03af8ec01c8332f766eb7b9c8f9.jpg"/>
  </ore:aggregation>

```

6 Content by partner

This report shows number of metadata records and digital objects by partner, as of 24th September 2014. The report is generated in the Event log, where it is updated in real time. The latest report is available at: http://locloudlog.avinet.no/report_content_by_partner.aspx.

Partner country	Partner name	Number of collections	Digital objects
AT	AIT	5	11100
BE	Provincie Limburg	2	60000
BG	PSRL	3	1000
CY	CUT	8	3905
CZ	NPU	2	6850
DE	UDE	8	
DK	KUAS	1	200000
ES	MECD	37	130000
FR	CG33	3	2416225
GR	Future Library	9	
HR	GKR	3	1100
IE	Discovery Programme	9	23000
IS	AHAI	2	1200
IT	FRS	6	5015
LT	VUKF	3	15000
NL	RCE	6	
NO	NRA	1	2020000
PL	PSNC	1	18300
PT	FMNF	6	
RO	BJC	1	1000
RS	BGB	2	1550
SE	ABMR	6	
SI	Zavod Jara	2	6564
SI	IPCHS	2	7000
SK	PrifUK KAEG	3	1000
TR	HU	1	1000
UK	UoY-ADS	6	19674
		total	4950483

Content lifecycle status by partner

The Event log is designed to track the activities involved in preparing content for harvesting and aggregation. This report summarises the events by partner. It is updated in real time as new events are entered into the system by partners. The latest report is available at:

http://locloudlog.avinet.no/report_collections_by_event2.aspx.

Partner name	No of collns	IPR cleared	Captured	Enriched	Published	PURIs	Exported	Harvested	Validated
AIT	5	4	3		4	3			
BJC	1	1							
BGB	2	0			1				
CG33	3	2	2		1	1	1		
CUT	8	0		8					
Discovery Programme	9	6	7		2	1			
FRS	6	6	2		2				
FMNF	6	2	3						
Future Library	9	0							
GKR	3	0							
HU	1	0	1						
PSNC	1	1	1		1	1	1		
Zavod Jara	2	2	2			1	1		
IPCHS	2	0				2			
KUAS	1	0	1		1	1			
RCE	6	0							
MECD	37	0							
AHAI	2	0							
NPU	2	1	1		1	1			
NRA	1	0							
PSRL	3	3	3		3		3		
Provincie Limburg	2	1			1		1		

Partner name	No of collns	IPR cleared	Captured	Enriched	Published	PURIs	Exported	Harvested	Validated
ABMR	6	6							
UDE	8	0							
UoY-ADS	6	6	6		6	6	6		
PrifUK KAEG	3	1		1					
VUKF	3	1	2						

7 Metadata quality

During the period between January to May 2014 a Europeana task force on metadata quality met to discuss how the quality of metadata supplied to Europeana could be improved and how this would improve end user experience. The expected outcomes of the task force were to:

- Document what Europeana means by metadata quality.
- Create best practice guidelines for creating good metadata quality records and good metadata quality fields.
- Make recommendations on metadata fields per media type.
- Provide documents to clarify the discrepancies between metadata displayed in the Europeana portal and provided data.

At the time of writing the Europeana Metadata Quality Task force had not published its report.

Europeana has a series of mandatory metadata elements, these are as follows:

- edm:dataProvider
- edm:isShownAt **or** edm:isShownBy
- edm:provider
- edm:rights
- dc:title **or** dc:description
- dc:language for text objects
- dc:subject **or** dc:type **or** dc:coverage **or** dc:spatial
- edm:type

The recommended metadata elements are:

- edm:object
- dcterms:alternative
- dc:creator
- dc:contributor
- dc:date
- dcterms:created
- dcterms:issued
- dcterms:temporal
- dc:publisher
- dc:source
- dcterms:isPartOf
- edm:isNextinSequence (for hierarchical objects)

One aspect of metadata quality is **completion** of the mandatory metadata elements. In fact records that do not fulfill the minimum mandatory requirements are automatically rejected by Europeana and will not be published in the portal.

There are other aspects of metadata quality which are routinely checked at the time of publication, these include:

- Uniqueness of identifiers
- Currency of links, i.e. check the links are not broken
- Coherence of the rights statement. Is the edm:rights statement (which is chosen from a standard list published by Europeana at: <http://pro.europeana.eu/available-rights-statements>) consistent with any rights statement included in the dc:rights element.

Other aspects of metadata quality, which can be checked for include:

- Use of international controlled vocabularies such as Geonames
- The length of titles and descriptive text
- Accuracy of geographical coordinates (for example, points lie within the expected region when mapped)

Length of text may not appear to be an aspect of metadata quality, however very short titles and descriptions offer limited information to end-users. They also limit the possibility of enriching the indexing of metadata records by using parsers to identify place names, personal names or other keywords.

The accuracy of geographic coordinates can be checked visually by plotting the coordinates onto a map. If a collection of heritage material from a northern European country is plotted somewhere in the South Pacific, there is a problem with the coordinates. Smaller discrepancies are less obvious, but can be checked and corrected using the LoCloud Geo microservice.

Apart from the availability of a reasonable description of an object, other aspects that Europeana has identified as being important for content re-use and having an impact on the quality of user experience include:

- The availability of a thumbnail or preview
- Provision of a direct link from the metadata record to an object (rather than to a landing page)
- Provision of higher resolution images
- Openly licencing which allows for re-use of the content that is made available

8 Conclusions

During the period under review LoCloud partners have made good progress with completing their content surveys.

Given the diversity of the content being provided by LoCloud partners it is difficult to identify standard metrics. However it is possible to monitor how the metadata compares to the recommendations of the Europeana quality task force, and to focus on improving those aspects that will have most impact on improving the representation of the content in Europeana, its potential for re-use and improving user experiences.

Statistical monitoring tools have been put in place to monitor the conformance of the metadata being provided to mandatory and recommended metadata elements. LoCloud microservices now implemented on the LoCloud aggregator can be used to address other aspects of metadata quality that improve its retrieval and usability.

Human monitoring of the metadata is also important. LoCloud partners are monitoring the metadata provided by their partners and working with them to improve its quality. In addition a sample of each collection is monitored by the project coordination, the staff of the LoCloud aggregation service and the content ingestion team at Europeana, and feedback is provided to relevant partner or content creator.