



*ECP-2007-DILI-517009*

## **EuropeanaLocal**

### **Performance measurement schema and end-user incentive schemes**

<b>Deliverable number</b>	<i>D-6.1</i>
<b>Dissemination level</b>	<i>Public</i>
<b>Delivery date</b>	<i>16 June 2009</i>
<b>Status</b>	<i>Final</i>
<b>Author(s)</b>	<i>Gunnar Urtegaard</i>



***eContentplus***

This project is funded under the *eContentplus* programme<sup>1</sup>,  
a multiannual Community programme to make digital content in Europe more accessible, usable and  
exploitable.

---

<sup>1</sup> OJ L 79, 24.3.2005, p. 1.

## **Contents**

<b>1</b>	<b>INTRODUCTION .....</b>	<b>3</b>
<b>2</b>	<b>AMOUNT, TYPES AND QUALITY OF CONTENT.....</b>	<b>4</b>
<b>3</b>	<b>WORKING GROUP MEETINGS IN VERIA AND ANCONA .....</b>	<b>5</b>
<b>4</b>	<b>LIFECYCLE OF A COLLECTION.....</b>	<b>6</b>
<b>5</b>	<b>ESTABLISHING THE EUROPEANALOCAL EVENT LOG.....</b>	<b>7</b>
<b>6</b>	<b>REPORTS FROM THE EVENT LOG .....</b>	<b>9</b>
<b>7</b>	<b>EVENT LOG AND PROGRESS REPORTS .....</b>	<b>10</b>
<b>8</b>	<b>METADATA QUALITY.....</b>	<b>11</b>
<b>9</b>	<b>CONCLUSIONS.....</b>	<b>12</b>

## 1 Introduction

EuropeanaLocal will put in place an infrastructure that will continue to increase the content available to Europeana. At the same time the project will enhance the skills, expertise and motivation required to support local institutions throughout Europe. Key content types to be made available through EuropeanaLocal include items and collections of high cultural value held at local or regional level, specific local collections held by libraries, museums and archives, local sound and film archives, public records held by archives, etc. EuropeanaLocal's role is thus to help create the conditions, through for example, training, advice, providing suitable mechanisms whereby local and regional institutions are in a position to contribute their content to Europeana.

As described in the Description of Work the objectives of WP6 are to:

1. Monitor and evaluate the amount, types and quality of metadata and content which are being contributed by EuropeanaLocal partners to Europeana according to a pre-defined set of targets and milestones
2. Analyse the extent to which the addition of local/regional content is contributing to the richness of content and value of services being delivered by Europeana (e.g. in identified thematic areas). Any areas of redundancy or duplication
3. Assess whether the impact of the work necessary and the way it is carried out during EuropeanaLocal constitutes a valid and achievable basis for the long-term growth, persistence and integration of local/regional content in Europeana.

This deliverable, D6.1, will focus on bullet point 1 i.e. how to monitor and evaluate the amount, types and quality of metadata and content being contributed from local and regional content providers. Bullet points 2 and 3 will be discussed and included in the overall evaluation framework in the Autumn of 2009. This is due to the close relationship between EuropeanaLocal and Europeana. Europeana have postponed the issue of thematic areas and EuropeanaLocal will address this issue at a later stage, following further clarification and direction from Europeana.

The Description of Work states that end users will be involved in the validation of the work of EuropeanaLocal through the establishment of user testing groups, consisting of demographically segmented groups (by age, level of education, social grouping ) of some 20-40 users in each partner region. This work will be closely co-ordinated with other user testing work being planned and carried out by Europeana. Not least because EuropeanaLocal has no separate portal for its content and users will access EuropeanaLocal content through the main Europeana portal. It will also be important to avoid duplication of work. For this reason the projects decided that it would be sensible to wait for the results of the recent Europeana user survey before carrying out any additional user testing. These results, due shortly, will be used to inform discussions between the two projects about what issues and aspects of the service might most benefit from further user testing and input. This in turn will influence what sort of user groups will be set up and for what particular purposes.

Bullet point three, impact study, will be an important topic at the next meeting on the Evaluation group in Lisbon in October 2009. This will also be carried out in liaison with Europeana ver. 1.0.

## 2 Amount, types and quality of content

An important goal for EuropeanaLocal is to increase significantly the quantity of high quality digital content which is accessible through Europeana. Broad quantitative targets and milestones are established and our headline performance indicators are:

Indicator Nr	Objective/expected result	Indicator name	Expected Progress		
			Year 1	Year 2	Year 3
1	Quantity of digital content from EuropeanaLocal accessible through Europeana	Content Items	0	3 million items	10 million items
2	Number of Europeana 'themes' agreed or generated with specific reference to EuropeanaLocal content	Themes	1	3	5
3	Number of EuropeanaLocal participant countries with repositories of local/regional content harvestable by Europeana (using Europeana metadata application profiles)	Harvestable Repositories	3	27	27
4	Number of new regional/local content providers accepted to join Europeana	New Content Providers	6	30	100
5	Number of countries with agreements to supply content to Europeana through nationally or regionally co-ordinated mechanisms	National Content Agreements	4	12	25
6	Proportion of use of Europeana which involves content supplied by EuropeanaLocal	Content Use	0	3 million items	10 million items

Expected progress in terms of amount and types of content should be 3 million items by end of year 2. This deliverable will describe how we intend to monitor progress through year 2 and further in year 3. No content is expected to be added in year 1.

### **3 Working group meetings in Veria and Ancona**

The main conclusion from these two meetings was to focus on how amount and types of metadata and content could be monitored. The group concluded that it is important to establish a service that will give access to updated information about progress in each country, each region, give information about participation from the three domains, and provide the possibility to trace each collection from local content provider to Europeana. The group also discussed how to evaluate quality.

The discussions in the group also stated that it is important to get key information equally from all partners/regions. The system established to collect this information must not be too difficult or involve too much work. The system must also give flexibility for those who wish to delve deeply into particular questions and share their thoughts with the rest of the project partners. The working group thus concluded that the monitoring and evaluation of the amount and types should be handled by a combination of manual reports and reports generated from local repositories extracting facts about exports, mapping of collections and so on.

The monitoring and evaluation work should follow each of the phases involved in the passage of a collection from a content provider to Europeana, and it should document the work done in each phase. This approach also creates a stronger link between the technical work done with functionality in the repositories and reporting facilities at collection and repository level. The group concluded that most of the data needed about amount and types could be collected by developing reports from the different phases in the lifecycle of a collection starting by entering data about collection in the content survey. The group introduced the term EuropeanaLocal event log.

The term event relates to the individual collection and is intended to document what is done with each digital collection relevant for EuropeanaLocal from the time the collection is entered into the content survey from partner or from new content providers. Data about each event will be added to an online event log. Some manual reports will also be needed to provide information about lessons learned and the identification of good practice. These reports can be uploaded to the event log and made available to other partners. Each partner will be responsible for their own reporting to the event log and also for reporting about new collections from new regions added later in the project.

Quality of metadata and content should be handled differently. Techniques will be developed by a combination of technical reports and manual report. A small group was established in Ancona to continue working on how to measure metadata quality. Once we have more metadata to analyze from different regions, the metadata quality group will be able to suggest some basic methods to analyze quality. These methods can be applied to individual collections or to repositories.

## 4 Lifecycle of a collection

The discussion in Veria and Ancona clearly stated that we should follow the lifecycle of a collection from EuropeanaLocal provider to Europeana and provide important data from each phase in the lifecycle.

Each collection being a candidate to Europeana through EuropeanaLocal **must be registered** in the content survey used for D2.1 Metadata survey report. The content survey contains important information about each collection. This online survey must be maintained and kept available for new content providers and new collections throughout the project. Through the survey important documentation about each collection are made available to the project and the evaluation group. The survey must represent all collections being identified as candidates for EuropeanaLocal.

There is no universal definition of the term collection, but in EuropeanaLocal it 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 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. This distinction is useful in the implementation process as more than one collection may be added to a local OAI-PMH compliant repository through one and the same metadata extraction/repository population operation. The workload of each content contributor partner largely depends on the number of different systems data have to be retrieved from.

The discussion in the working group identified the following activities as important milestones to document at collection level:

1. At the point when the collection is added to the content survey.
2. At the point when metadata is extracted about all items in a collection and mapped to the ESE. (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).
3. At the point when metadata is normalized. (i.e. the process of transforming attribute values from one notation to another. E.g. a standardized way of expressing dates, transformation of coordinates etc. This may apply to some collections or only some items in each collection, but may be not to all).
4. At the point when metadata is enriched. (i.e. automatically or semi-automatically processing of metadata with the purpose of improving the quality of what, who, where and when metadata. This may also apply to some collections or items, but may be not to all).
5. At the point when collections are populated into a regional or national repository. (i.e. the process of loading data into the chosen repository software based on custom metadata extracts or exports of known standard formats).
6. At the point when collection are harvested from regional/national repository by Europeana. (i.e. the process of connecting to a repository, issuing a request for data and downloading metadata content as XML).
7. At the point when the collection is added to Europeana services and is available to end users.

## 5 Establishing the EuropeanaLocal Event log

The event log suggested by the evaluation group will be a simple, but powerful tool to monitor amount and progress. At the same time the event log will be open to all partners and content providers as a place to add comments on framework, tools, methods and standards being involved in the project. Let us use an example.

In the content survey a photo collection of 70.000 items and objects from Sogn og Fjordane, Norway, is listed. The next step is that the collection's metadata is extracted by the aggregator and mapped to DC ESE 3.1. The last important step is when the metadata is added to EuropeanaLocal and included in the services.

The content survey contains the important information about the collection, like amount of items, objects, thumbnails, metadata formats, technical formats etc. The event log will collect and document work done on this particular collection from the original content provider to Europeana and any problems, questions and considerations done by the provider or the aggregator. They may add manual reports to each event about technical issues, the use of ESE 3.1., normalization problems etc.

Country	Region	Provider			
Norway	Sogn og Fjordane	County Archives in Sogn og Fjordane			
Themes	Historical photographs depicting persons, rural life, rural landscapes, industry, social life, life cycle rituals, farming, buildings, emigration,				
Timespan	1860-2008	Items: 70000			Objects: 70000
<b>Event log:</b>					
	Date added	Items	Objects	Added by	Report
<b>Content survey</b>	21.5.2009	70000	70000	NN	URL
<b>Extraction</b>	22.5.2009	70000	55000	NN	URL
<b>Mapping</b>	21.5.2009	70000		NN	URL
<b>Normalization</b>	21.6.2009	10000		NN	URL
<b>Enrichment</b>	21.6.2009	10000		NN	URL
<b>Repository population</b>	25.6.2009	70000	55000	NN	URL
<b>Metadata harvested</b>	25.7.2009	70000	55000	NN	URL
<b>Collection part EuropeanaLocal service</b>	30.7.2009	70000	55000	NN	URL

**Comments:**

The Event log will be a simple database with suitable reporting and analyzing functions. The services will be available from the EuropeanaLocal web site. When a content provider or aggregator has done work on one or several collections, he/she must connect to the event log, select the collection involved and add information. 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.

The aggregator of this photo collection from Norway will connect to the event log, search for this particular collection and at the simplest just add data about extraction and mapping, and the number of items and objects effected. Mapping, normalization and enrichment will only affect metadata, and thus different numbers will be reported. Country, region, provider, themes, timespan, amount and items and objects will be selected automatically from the content survey by using the collection identifier.

Information to be added about each event:

<b>Info</b>	<b>Comments</b>
Type of event	Selected from listbox
Date	Automatic
Number of items handled	Normalization, mapping and enrichments only effects metadata and items, and a selection of items
Number of objects	Only effects extraction and harvesting
Name of person responsible	Elin Østevik
Email of person responsible	Elin.ostevik@sfj.no
Issues related to the event	Short comments
Lessons learned	Comments
Uploaded documents.	Manual reports and documentation

The technical partner, Avinet, will set up the event log as a prototype by July 2009 and as a fully functional service by September 1. 2009 in time for monitoring the amount of data transported from the local and regional providers to Europeana. The connection between the content survey and the event log will be maintained through a persistent collection identifier. A handful of partners responsible for the first collections to be harvested will test the event log and the results of these tests will be discussed in the next meeting of the evaluation Working group in October 2009.



**EuropeanaLocal Event Submission Form**

Type of event: Harvesting

Date: 15.07.2009

Number of items handled: n/a

Number of objects: 50000

Name of person responsible: Antonio Vieira

E-mail of person responsible: antonio@vieira.pt

Issues related to the event: First attempt failed due to network time-out issues

Lessons learned: Must time the harvesting so that it doesn't coincide with other types of heavy network usage.

File for upload: C:\report.xml

Buttons: Browse..., Cancel, Submit Event

Fig 1: Example of event submission form

## 6 Reports from the event log

The event log will be a very important tool for collecting and sharing information about the amount of content and progress in EuropeanaLocal. Reports can be selected by country, region, type of content, time span, amount etc. The number of collections will not be very large, even at the end of the project.

The event log will document what happened to each collection. Search and reporting facilities available for all partners will be developed. The event log will enable many interesting analyses of the data to be performed, for example:

- The event log will show at any given time the amount of items and objects being extracted from local providers and the amount added to Europeana. This will be a report organized by country and updated on a daily basis by the system collected data from the event log
- How many collections, items and objects are entered into the content survey at any given time? (Can easily be arranged by country, domain etc)
- How many digital objects from a chosen number of collections are extracted and mapped to Europeana by a certain data.
- 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 item and objects are harvested into 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 EuropeanaLocal at any given time.
- The number of manual reports added, from which country, provider, etc

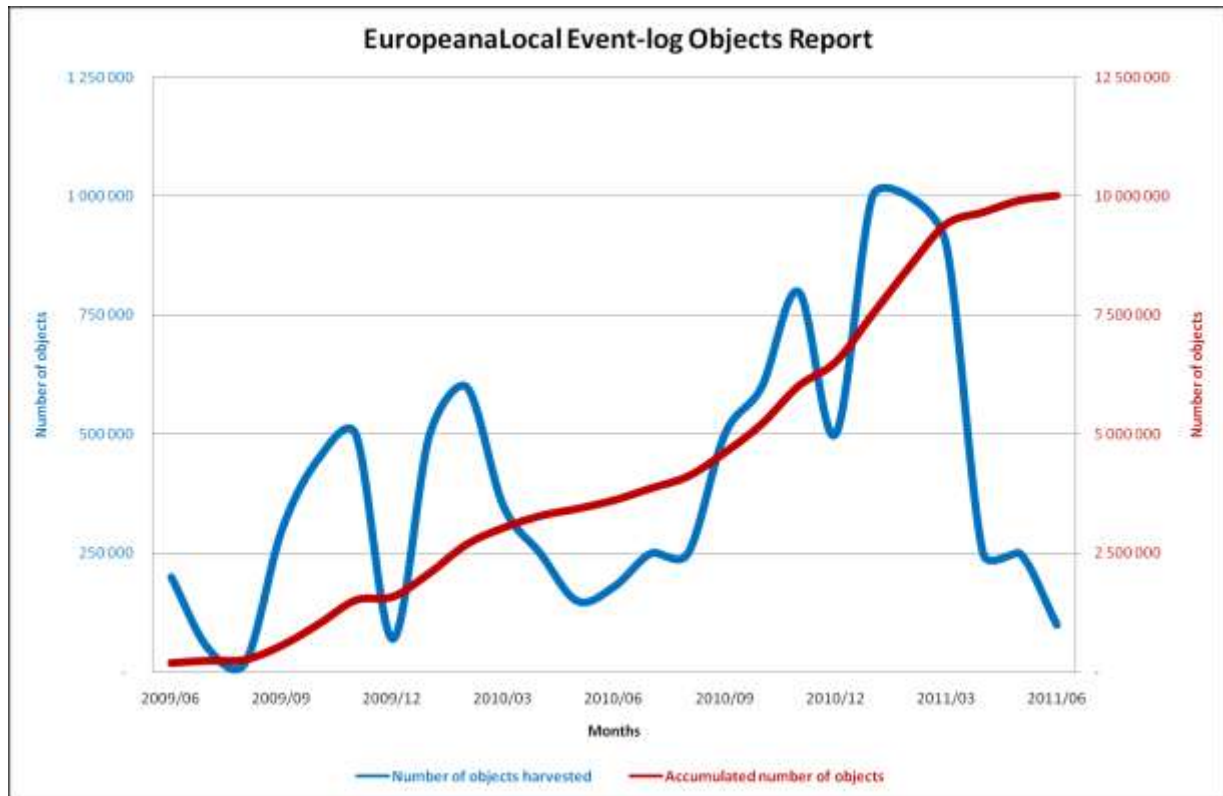


Fig 2: Example of possible output from EuropeanaLocal event log

## 7 Event log and progress reports

The Description of Work indicates that a progress report (D6.2) should be delivered every 6 months showing progress based on a predefined set of key metrics. The event log will be the main tool for creating data for these reports. Data will be downloaded from the event log and analyzed and results reported in D6.2. The manual reports added to the event log will also constitute an important resource to be used in these progress reports.

As described in the Description of Work another of the objectives of WP6 is to:

*Assess whether the impact of the work necessary and the way it is carried out during EuropeanaLocal constitutes a valid and achievable basis for the long-term growth, persistence and integration of local/regional content in Europeana.*

The manual reports added to the event log will give us access to experiences collected by providers and partners and aggregators throughout the project.

### Information to the partners about the event log.

By September 2009, when the event log is established, a manual will be ready and distributed to all the partners giving them information about the event log and how to use it as a tool to document own actions and get access to lessons learned by others.

This manual will describe in detail how to use the event log, why it is established and how it will benefit the project and document the work done.

## 8 Metadata quality

The Evaluation Working group discussed the issue of metadata quality. What is good metadata and how can it be measured? It is important that the metadata that contributes to the EuropeanaLocal should be of as high a quality as possible. Thus, we should have a way to measure their quality. Metadata with higher quality will have an extra incentive to be included in the EuropeanaLocal collections. The diversity of the metadata, and the way they will be used, makes it hard to compare quality among them, and to use a global metric for that. For example, different criteria apply to free-text metadata than apply to controlled vocabularies. The first 209 institutions that are participating in EuropeanaLocal are using the following metadata schemes:

Dublin Core	127	59.9%
ISAD(G)	62	29.2%
METS	53	25.0%
EAD	49	23.1%
MARC	39	18.4%
SPECTRUM	15	7.1%
MAB	8	3.8%
TEI	7	3.3%
MODS	4	1.9%
Object ID	5	2.4%
CIDOC-CRM	2	0.9%
VRA	1	0.5%
Museumdat	1	0.5%
CDWA	1	0.5%

Some of them have also adapted the metadata standards that they use. Each metadata schema would normally need its own metadata evaluation metrics. Since OAI will gather all metadata in Dublin Core with the Europeana Data Set Extensions, the evaluation of the quality of this outcome must be measured in the collected metadata.

Similar studies<sup>1</sup> on NSDL, also an OAI harvested metadata, measure the quality of the metadata of each collection by the number of Dublin Core fields they use. In addition to doing that, we can also assign appropriate weights to the existence or the repetition of the different

---

<sup>1</sup> e.g. Bui, Yen, Park, Jung-ran (2005). *An assessment of metadata quality: a case study of the National Science Digital Library Metadata Repository*. In Haidar Moukdad (Ed.) CAIS/ACSI 2006 Information Science Revisited: Approaches to Innovation. Proceedings of the 2005 annual conference of the Canadian Association for Information Science held with the Congress of the Social Sciences and Humanities of Canada at York University, Toronto, Ontario. [Access at 19/3/2009] <[http://www.cais-acsi.ca/proceedings/2006/bui\\_2006.pdf](http://www.cais-acsi.ca/proceedings/2006/bui_2006.pdf)>

fields, according to their significance. A small group led by the Greek partner will continue to work on this issue. The group will discuss relevant tools and methods that will take as input the harvested metadata of a collection, and will produce some metrics on its quality, based on the existence and the repetitions of the Dublin Core fields in it. The same tools and methods may also be applied to all EuropeanaLocal harvested collections, and possibly to some Europeana collections that we will use as reference collections for comparison purposes. The group will analyse the results and produce a report based on the results, focusing mostly on what metadata ideally ought to be present in new collections in order to produce the best results. Metadata quality will be an important topic in the next meeting in Lisbon Autumn 2009.

## 9 Conclusions

This deliverable summarises the work of the Evaluation Working Group to date. The group is made up of representatives from the following partners: Sogn og Fjordane County Municipality, MDR Partners, Roskilde Kommune, Veria Central Public Library, Regione Marche, EEA, Asplan Viak, ABM-Utvikling, Fundacao Museu Nacional Ferroviario. ABM-utvikling is the work package leader.

The work presented here focuses on the creation of an Event Log which will be used to monitor and evaluate the amount, types and quality of metadata and content which are being contributed by EuropeanaLocal partners to Europeana. The log will be used to collect evaluative data at key points in the content lifecycle as follows:

1. At the point when the collection is added to the content survey.
2. At the point when metadata is extracted about all items in a collection and mapped to the ESE. (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).
3. At the point when metadata is normalized. (i.e. the process of transforming attribute values from one notation to another. E.g. a standardized way of expressing dates, transformation of coordinates etc. This may apply to some collections or only some items in each collection, but may be not to all).
4. At the point when metadata is enriched. (i.e. automatically or semi-automatically processing of metadata with the purpose of improving the quality of what, who, where and when metadata. This may also apply to some collections or items, but may be not to all).
5. At the point when collections are populated into a regional or national repository. (i.e. the process of loading data into the chosen repository software based on custom metadata extracts or exports of known standard formats).
6. At the point when collection are harvested from regional/national repository by Europeana. (i.e. the process of connecting to a repository, issuing a request for data and downloading metadata content as XML).
7. At the point when the collection is added to Europeana services and is available to end users.

This data and the reports generated through the events log will provide an invaluable source of user feedback to Europeana and other interested parties, which can be used to inform both

modifications and further developments to the systems and processes they make available to content providers.

Over the next period the Group will review the operation of the Event Log, based on practical experience from its use as data begins to be harvested from the first group of partners. They will use the data from the log to complete the first 6 monthly evaluation report (D 6.2) and will consider how best to conduct the impact study.