*ECP-2007-DILI-517009*

**Europeana Local**

# Deliverable 3.2: Technical report on initial harvesting results

| | |
|---|---|
| **Deliverable number** | *D3.2* |
| **Dissemination level** | *PU; Public* |
| **Delivery date** | *February 2010* |
| **Status** | *Final* |
| **Author(s)** | *Valentine Charles, EDL Foundation* |

*e***Content***plus*

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

## Contents

## 1   Introduction

EuropeanaLocal has the task of making available digital content from local and regional
institutions through Europeana. To make this process manageable this involves the
establishment of a harvestable network of OAI-PMH compliant metadata repositories to
aggregate this content.

In summer 2010 the operational service of Europeana will be released, with the aim of ten
million digital items accessible online. EuropeanaLocal plays an important role in reaching
this aim by providing a large amount of content via OAI-PMH and testing data in the
Europeana Content Checker, a test environment provided by the EDL Foundation.

---

[1]  OJ L 79, 24.3.2005, p. 1.

## 2   Scope

This document relates to *Milestone 2: successful completion of test harvesting procedures (usable content/metadata from six countries indexed in central environment)  (M18)*, as set out in Workpackage 3 – Content harvesting and aggregation.

More specifically it relates to:

- *Task 2.5: Test harvesting, in batches of 3 countries per month, will be carried out on the installed repositories of local/regional content by the EDL Foundation.  Any local modifications required will be carried out with the assistance of the project-wide technical support partners as necessary. Any modifications required to the central metadata environment will be carried out by EDL Foundation. Work to be completed for all EuropeanaLocal countries by Month 30.  Feedback will inform the modification of tools and services by EDL Foundation.*

**Deliverable 3.2 Technical report on intitial harvesting results** summarises the results of the harvesting tests that were carried out on  the installed repositories of local/regional content by the EDL Foundation. Data from local/regional content was first tested by six EuropeanaLocal content providing partners in the Europeana Content Checker, before submitting the repository URL to the Europeana office. EDLF ran harvesting tests on the data from these six[1] EuropeanaLocal partners, as described in Task 2.5 to be able to make any nesserary changes. EDL Foundation gave its feedbacks to the partners through these test harvesting results to enhance the quality of tools and data.

## 3   Harvesting tests

Harvesting tests were carried out on repositories of local/ regional institutions from EuropeanaLocal. The Europeana office tests functionalities that are defined in the OAI-PMH (The Open Archive Initiative Protocol for Metadata Harvesting) specifications and answered to the Europeana requirements. It also tests the compatibility with the Europeana harvesting client. The aim of these tests is to verify that the local repository is compliant with Europeana requirements.  Tests are carried out until the status indicate that data are ready to be harvested in production.

### 3.1   Results of harvesting tests

Six harvesting tests have been done and after evaluation, data were ingested in production. EDL Foundation did not detect any errors in the way partners from EuropeanaLocal had implemented their OAI-PMH repositories. Local/regional repositories are supporting the

---

[1]   Federacja Bibliotek Cyfrowych(PL), ABM-Utvikling(NO), CultureGrid (UK), Angewandte Informationstechnik Forschungsgese (AT), EuropeanaLocal- Zentral- und Landesbibliothek Berlin (DE), Ministry of Spain (ES).

Europeana Semantic element (ESE) schema, and requirements defined by the OAI-PMH protocol that are mandatory for Europeana Local.
Results of these harvesting test are provided in Annexes A1-A6.

### 3.2 Evaluation of data after harvesting

Partners from EuropeanaLocal have tested their data in the Content Checker, before submission of it for harvesting. That's why the evaluation of data after harvesting by the Europeana office gave good results. Only a few and not critical errors have been noticed like spelling mistake, encoding character issues, problems with thumbnails. These errors have been immediately fixed by the EuropeanaLocal content providers, data has been re-tested and re-submitted.

## 4 Conclusion

This report summarises the current status of the content ingestion from EuropeanaLocal to Europeana. Europeana Guidelines for Mapping & Normalisation and the access to the Content Checker provides support to partners in the work and allows them to deliver quality data to Europeana. Results of this deliverable proves that the workflow between EuropeanaLocal and Europeana is adapted and allows on-going procedure.

At the moment harvested and approved EuropeanaLocal content from six partners has been indexed and is available via Europeana.eu, disclosing the great diversity of cultural heritage held at local and regional level throughout Europe.

# Annexe A- Europeana Production Ingestion-Acceptance tests for OAI-PMH harvesting.

## *Annexe A1: EuropeanaLocal - Federacja Bibliotek Cyfrowych (PL)*

### GENERAL INFORMATION

**Date**

2009-11-27

**Tester name and contact info**

Julie Verleyen, julie.verleyen@kb.nl

**Provider Name**

EuropeanaLocal - Federacja Bibliotek Cyfrowych

**Provider ID**

09404

**OAI-PMH repository base URL**

http://fbc.pionier.net.pl/owoc/op-metadata

### TEST RESULTS

**Note:** The tests focus on the key OAI-PMH features that are required for harvesting by Europeana in the Production environment.

**Support of 6 request verbs: `Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord`**

OK

**Support of `metadataPrefix` parameter: 'ese'**

OK

**Support of `resumptionToken` attributes: `expirationDate, completeListSize, cursor`**

OK

**Harvesting specification for Europeana of `setSpec` to be used by Europeana:**

> **Base URL**
> http://fbc.pionier.net.pl/owoc/op-metadata
>
> **setSpec**
> europeana

**Example**

http://fbc.pionier.net.pl/owoc/op-metadata?verb=ListRecords&metadataPrefix=ese&set=europeana

**Compatibility with Europeana harvesting client**

OK

**Remarks/recommendations**

None

## *Annexe A2: EuropeanaLocal-ABM-Uitvikling (NO)*

### GENERAL INFORMATION

**Date**

2009-11-27

**Tester name and contact info**

Julie Verleyen, julie.verleyen@kb.nl

**Provider Name**

EuropeanaLocal – ABM-Utvikling

**Provider ID**

09414

**OAI-PMH repository base URL**

http://93.94.14.7/harvester2/index.php/oai

http://oaipmh.dimutest.kulturit.no/

**TEST RESULTS** - http://93.94.14.7/harvester2/index.php/oai

**Note:** The tests focus on the key OAI-PMH features that are required for harvesting by Europeana in the Production environment.

**Support of 6 request verbs: `Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord`**

Problem with following:
http://93.94.14.7/harvester2/index.php/oai?verb=GetRecord&identifier=oai:pkp.no:record/712730&metadataPrefix=ese

**Support of `metadataPrefix` parameter: 'ese'**

OK

**Support of `resumptionToken` attributes: `expirationDate, completeListSize, cursor`**

OK

**Harvesting specification for Europeana of `setSpec` to be used by Europeana:**

 **Base URL**
 http://93.94.14.7/harvester2/index.php/oai?

 **setSpec**
 sffFoto
 sffKL
 sffDF

 **Example**
Photos (55008 records):
http://93.94.14.7/harvester2/index.php/oai?verb=ListRecords&metadataPrefix=ese&set=sffFoto
#Articles (1621):
http://93.94.14.7/harvester2/index.php/oai?verb=ListRecords&metadataPrefix=ese&set=sffKL
#Articles some including audio/video (843):
http://93.94.14.7/harvester2/index.php/oai?verb=ListRecords&metadataPrefix=ese&set=sffDF

**Compatibility with Europeana harvesting client**

OK

**Remarks/recommendations**

No support for deleted records: this means that there is no automatic way for Europeana to be in sync with the source repository if records are deleted from the latter. In this situation ABM-Utvikling will need to specify to Europeana if records have been deleted.
It is recommended that ABM-Utvikling indicates how deleted records synchronization with Europeana can be dealt with.

| | |
|---|---|
| **Formatted:** Italian (Italy) | |
| **Field Code Changed** | |
| **Formatted:** Italian (Italy) | |
| **Formatted:** Italian (Italy) | |

**TEST RESULTS - http://oaipmh.dimutest.kulturit.no/**

**Note:** The tests focus on the key OAI-PMH features that are required for harvesting by Europeana in the Production environment.

**Support of 6 request verbs: `Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord`**

OK with a remark regarding GetRecord:

<span style="color:red">Remark regarding GetRecord</span>

A `GetRecord` request with the `identifier` parameter doesn't work because the identifier of each record includes an ampercent sign ('&') which can be interpreted as a separator for URL parameters.

Example:
A GetRecord request for identifier *oai:digitaltmuseum.no:idOwner=SFM&idIdentifier=DHS.00304* is as follows:
http://oaipmh.dimutest.kulturit.no/?verb=GetRecords&identifier=oai:digitaltmuseum.no:idOwner=SFM&idIdentifier=DHS.00304&metadataPrefix=ese

It looks as if *idIdentifier* is a URL parameter but this is an unknown parameter for the OAI-PMH protocol.

The current Europeana harvesting procedure doesn't make GetRecord calls to the providers' repositories so this is not an issue for now. However we recommend that values of parameters don't include parameters that

**Support of `metadataPrefix` parameter: 'ese'**

OK

**Support of `resumptionToken` attributes: `expirationDate, completeListSize, cursor`**

OK

**Harvesting specification for Europeana of `setSpec` to be used by Europeana:**

> **Base URL**
> http://oaipmh.dimutest.kulturit.no/

> **setSpec**
> SFM

> **Example**
http://oaipmh.dimutest.kulturit.no/?verb=ListRecords&metadataPrefix=ese&set=SFM

**Compatibility with Europeana harvesting client**

OK

**Remarks/recommendations**

*Remark*

No support for deleted records: this means that there is no automatic way for Europeana to be in sync with the source repository if records are deleted from the latter. In this situation ABM-Utvikling will need to specify to Europeana if records have been deleted.
It is recommended that ABM-Utvikling indicates how deleted records synchronization with Europeana can be dealt with.

***Recommendation***
Regarding syntax of OAI-PMH record identifier: Avoid characters that can be interpreted as parameters separators in the URL. Ex: '&' or '='

## Annexe A3: EuropeanaLocal-CultureGrid (UK)

**TEST CONCLUSION**

**OK for harvesting in Production**

**Please provide support of `resumptionToken` attributes: `expirationDate`, `completeListSize, cursor`**

**GENERAL INFORMATION**

**Date**

2009-12-01

**Tester name and contact info**

Julie Verleyen, julie.verleyen@kb.nl

**Provider Name**

EuropeanaLocal – CultureGrid

**Provider ID**

09405

**OAI-PMH repository base URL**

http://culture-grid.k-int.com/dpp/oai

**TEST RESULTS**

- **Support of 6 request verbs: `Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord`**

OK

- **Support of `metadataPrefix` parameter: 'ese'**

OK

- **Support of `resumptionToken` attributes: `expirationDate, completeListSize, cursor`**

NOT OK

- **Harvesting specification for Europeana:**

  **Base URL**
  http://culture-grid.k-int.com/dpp/oai

  **setSpec**
  PN:BeamishTreasures
  PN:BedesWorld
  PN:BowesOAI
  PN:DurhamCountyCouncil
  PN:EnglishHeritage
  PN:iSeeGateshead
  PN:KirkleesImages
  PN:LambethLandmark
  PN:Leodis
  PN:PictureThePast
  PN:PortableAntiquities
  PN:TWImagine

  **Example**
http://culture-grid.k-int.com/dpp/oai?verb=ListRecords&set=PN:BeamishTreasures&metadataPrefix=ese

- **Compatibility with Europeana harvesting client**

OK

- **Remarks/recommendations**

None

## *Annexe A4: EuropeanaLocal – Angewandte Informationstechnik Forschungsgese (AT)*

**TEST CONCLUSION**

**OK for harvesting in production**

**GENERAL INFORMATION**

**Date**

2010-01-06

**Tester name and contact info**

Julie Verleyen julie.verleyen@kb.nl

**Provider Name**

EuropeanaLocal – Angewandte Informationstechnik Forschungsgese

**Provider ID**

09403

**OAI-PMH repository base URL**

http://www.europeana-local.at/mdm/oai/

## TEST RESULTS

**Note:** The tests focus on the key OAI-PMH features that are required for harvesting by Europeana in the Production environment.

- **Support of 6 request verbs: `Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord`**

OK

- **Support of `metadataPrefix` parameter: 'ese'**

OK

- **Support of `resumptionToken` attributes: `expirationDate, completeListSize, cursor`**

OK

- **Harvesting specification for Europeana:**

    **Base URL**
    http://www.europeana-local.at/mdm/oai/

    **setSpec**
    Records are organized by set.
    Europeana will harvest the full repository without specifying any set. In other terms: Europeana assumes that the sum of the records of all sets corresponds to the full aggregated data.

    **Example**
http://www.europeana-local.at/mdm/oai/?verb=ListRecords&metadataPrefix=ese

- **Compatibility with Europeana harvesting client**

OK

- **Remarks/recommendations**

None

## Annexe A5: EuropeanaLocal- Zentral- und Landesbibliothek Berlin (DE)

**TEST CONCLUSION**

*OK for harvesting in production*

*Please address the Remarks/recommendations section and report back to Europeana.*

**GENERAL INFORMATION**

**Date**

2010-01-29

**Tester name and contact info**

Valentine Charles: valentine.charles@kb.nl

**Provider Name**

EuropeanaLocal- Zentral- und Landesbibliothek Berlin

**Provider ID**

09428

**OAI-PMH repository base URL**

http://130.73.102.69/eld/oai/harvester2/index.php/oai

## TEST RESULTS

**Note:** The tests focus on the key OAI-PMH features that are required for harvesting by Europeana in the Production environment.

- **Support of 6 request verbs: `Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord`**

OK with a remark regarding GetRecord and ListIdentifiers

ListIdentifiers is not supported for the metadataPrefix=ese
http://130.73.102.69/eld/oai/harvester2/index.php/oai?verb=ListIdentifiers&metadataPrefix=ese
GetRecord is not working
http://130.73.102.69/eld/oai/harvester2/index.php/oai?verb=GetRecord&identifier=oai::record/33122&metadataPrefix=ese

- **Support of `metadataPrefix` parameter: 'ese'**

OK

- **Support of `resumptionToken` attributes: `expirationDate, completeListSize, cursor`**

OK

- **Support of deleteRecords: no, transient, persistent**

OK

- **Harvesting specification for Europeana:**

  **Base URL**
  http://130.73.102.69/eld/oai/harvester2/index.php/oai

  **setSpec**
  Records are organized by set.
  Europeana will harvest the full repository without specifying any set. In other terms: Europeana assumes that the sum of the records of all sets corresponds to the full aggregated data.

  **Example**
  http://130.73.102.69/eld/oai/harvester2/index.php/oai?verb=ListRecords&metadataPrefix=ese

- **Compatibility with Europeana harvesting client**

OK

- **Remarks/recommendations**

None

## Annexe A6: Europeana-Local Ministry of Spain (ES)

| TEST CONCLUSION |
| --- |
| *OK for harvesting in production* |

| GENERAL INFORMATION |
| --- |

**Date**

2010-02-12

**Tester name and contact info**

Wibke Kolbmann  wibke.kolbmann@kb.nl

**Provider Name**

Europeana-Local Ministry of Spain

**Provider ID**

09407

**OAI-PMH repository base URL**

http://www.digibis.com/oai_ese/i18n/oai/oai.cmd

## TEST RESULTS

**Note:** The tests focus on the key OAI-PMH features that are required for harvesting by Europeana in the Production environment.

- **Support of 6 request verbs: `Identify, ListSets, ListMetadataFormats, ListIdentifiers, ListRecords, GetRecord`**

http://www.digibis.com/oai_ese/i18n/oai/oai.cmd?verb=Identify  Ok
http://www.digibis.com/oai_ese/i18n/oai/oai.cmd?verb=ListSets OK
http://www.digibis.com/oai_ese/i18n/oai/oai.cmd?verb=ListMetadataFormats        OK
http://www.digibis.com/oai_ese/i18n/oai/oai.cmd?verb=ListIdentifiers&metadataPrefix=ese OK
http://www.digibis.com/oai_ese/i18n/oai/oai.cmd?verb=ListRecords&metadataPrefix=ese&set=ese OK
http://www.digibis.com/oai_ese/i18n/oai/oai.cmd?verb=GetRecord&identifier=oai:hispana.mcu.es:9748&metadataPrefix=ese OK

- **Support of `metadataPrefix` parameter: 'ese'**

OK

- **Support of `resumptionToken` attributes: `expirationDate, completeListSize, cursor`**

OK
<resumptionToken    expirationDate="**2010-02-12T11:05:27Z**"    completeListSize="**849827**" cursor="**0**">**1265969127463**</resumptionToken>

- **Support of deleteRecords: no, transient, persistent**

OK

  <deletedRecord>**persistent**</deletedRecord>

- **Harvesting specification for Europeana:**

  **Base URL**

http://www.digibis.com/oai_ese/i18n/oai/oai.cmd

  **setSpec**
  *ese*

  **Example**

http://www.digibis.com/oai_ese/i18n/oai/oai.cmd?verb=ListRecords&metadataPrefix=ese&set=ese

- **Compatibility with Europeana harvesting client**

OK

- **Remarks/recommendations**

No remarks