Deliverable

Project Acronym: thinkMOTION
Grant Agreement number: 250485
Project Title: Digital Mechanism and Gear Library goes Europeana

D2.1 - Intermediate report on DMG-Lib OAI-PMH interface
Revision: 1.1

Authors:
Ulf Döring (Ilmenau University of Technology)

---

<p>| Project co-funded by the European Commission within the ICT Policy Support Programme |</p>
<table>
<thead>
<tr>
<th>Dissemination Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
</tr>
<tr>
<td>C</td>
</tr>
</tbody>
</table>
## Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Author</th>
<th>Organisation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>25.04.2012</td>
<td>U. Döring</td>
<td>IUT</td>
<td>Draft</td>
</tr>
<tr>
<td>0.2</td>
<td>22.06.2012</td>
<td>U. Döring</td>
<td>IUT</td>
<td>Draft</td>
</tr>
<tr>
<td>0.3</td>
<td>25.06.2012</td>
<td>U. Döring</td>
<td>IUT</td>
<td>Draft, latest results</td>
</tr>
<tr>
<td>0.4</td>
<td>28.06.2012</td>
<td>S. Falke</td>
<td>IUT</td>
<td>Review</td>
</tr>
<tr>
<td>1.0</td>
<td>22.11.2012</td>
<td>U. Döring</td>
<td>IUT</td>
<td>Final version</td>
</tr>
<tr>
<td>1.1</td>
<td>23.11.2012</td>
<td>S. Falke</td>
<td>IUT</td>
<td>Review</td>
</tr>
</tbody>
</table>

## 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This deliverable contains original unpublished work except where clearly indicated otherwise.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. Workflow and results.............................................................................................................................. 4
3. Technical details concerning ESE support in the DMG-Lib OAI-PMH interface ................. 7
   ESE 3.4 conformance .............................................................................................................................. 7
   Data Quality checking ............................................................................................................................ 7
   Fast adaptability of functionality ........................................................................................................... 7
   Caching of Responses ............................................................................................................................ 7
   Debug code .............................................................................................................................................. 7
4. Results ................................................................................................................................................ 8
ANNEX I – Well-structured XML-data for automatic harvesting of items by Europeana ............... 9
1. Introduction
In June 2010 partners from six European universities started the project thinkMOTION with the main objective of providing content from the field of motion systems via the Europeana online portal. Therefore, the currently available and all the DMG-Lib content collected in the proposed project must become accessible to Europeana - that means to Europeana users as well as to Europeana tools (e.g. the harvester).

Europeana must be able to create an internal representation of DMG-Lib’s object space. For the data transfer the existing OAI-PMH interface is extended with ESE (Europeana Semantic Elements) support, which is part of task 2.1 and described in this document.

Task 2.1 is part of WP2 – Adaptation of interfaces to Europeana – included in the “Description of Work” for thinkMOTION project. WP2 is the basis for the final step in WP5 (Task 5.5), where the content is integrated into the online portal of DMG-Lib and made accessible for Europeana.

2. Workflow and results
The existing OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) interface was extended to deliver metadata and thumbnails in an ESE (Europeana Semantic Elements) conform way. This was connected with the necessary formal “Steps to contribute directly to Europeana” (Figure 1).

Figure 1. Steps to contribute directly to Europeana

START

1. Organisation + Europeana Office sign Data Provider Agreement or Data Aggregator Agreement

2. Organisation submits the Submission Form to Europeana Office

3. Europeana Office processes the Submission Form within 1 month

4. Organisation prepares and validates data and transfer method according to Submission Form

5. Europeana Office validates transfer method

6. Europeana Office ingests the Organisation’s data according to the agreed schedule and notifies by e-mail the Organisation when the ingestion process is completed

END
Activities were as follows:

• **Sign the Data Exchange Agreement (DEA)** which is available since September 2011. We signed the agreement on November 7th, 2011 and sent it to Europeana office where these DEAs were collected for later signing by Europeana (Figure 2).

• **Get an ID** by Europeana for thinkMOTION (20208) and exchange contact details, especially for the access to the Content Checker.

• **Checking generated ESE-records in the Content Checker.** Finally, the metadata integrity checks were successful in May and begin of June 2012 (Figure 3). The content was shown in the Europeana test portal where e.g. the transferred items are grouped by type and the links back to DMG-Lib worked.

• **Send the submission form** to the ingestion team. After passing the tests with the Content Checker the form was sent on June 6th, 2012.

• **Give support in adjusting the metadata transfer** from DMG-Lib to Europeana. After some clarification mails concerning the information in the submission form, especially concerning the possibilities to harvest based on sets, the ingestion team started to access the OAI-PMH interface of www.dmg-lib.org. In June 2012 we were in tide contact with the ingestion team, esp. Dimitra Atsidis, for finding the best metadata transfer method and an Europeana-conform design of the metadata records. To improve the probability for that, different access ways were supplied to the ingestion team: standard OAI-PMH implementation, cached OAI-PMH access for a very fast response, direct link to an XML-file with the metadata for fastest access. Finally, the first DMG-Lib content was successfully ingested into Europeana at the beginning of July 2012 (see Figure 4).

• **Enabling automatic harvesting of DMG-Lib items.** The first ingestions had to be done by hand. Unfortunately, the Europeana tools stopped the automatic ingestion process due to suspicious errors, which gave no hints about the real reason for ingestion problems. During a Europeana Project Group Meeting in Den Haag, a one-to-one talk and debugging session with a developer from the Europeana team unveiled, that the problems were caused by the XML-data structure. After re-structuring the XML-data, the Europeana tools successfully harvested DMG-Lib items (see ANNEX I – Well-structured XML-data for automatic harvesting of items by Europeana).

Figure 2. Screenshot of list of signees of the Europeana Data Exchange Agreement
Figure 3. Screenshot of successful metadata integrity check in the Europeana test portal

Figure 4. Screenshot of successful ingestion of DMG-Lib content into Europeana
3. Technical details concerning ESE support in the DMG-Lib OAI-PMH interface

ESE 3.4 conformance
The Europeana Semantic Elements Specification defines a set of rules concerning the existence of certain metadata fields in the records sent as response to Europeana harvesting requests. The existing OAI-interface classes were extended to allow checking if a certain item is ESE conform or not. Thus an item record may be sent via OAI_DC-format but because of incompatibility with ESE specification it is not sent in ESE-format until the missed data was entered.

Data Quality checking
The experiences from the work in the last two years have shown that for some metadata fields formal quality checks can be performed to give the team members a fast feedback about the current quality state of a certain item. Therefore two icons are integrated which show the current state of an item in the project internal production database (ProDB). For instance shows that the item will be shown in the DMG-Lib portal but is not sent to Europeana – may be because of missed ESE conformance or because of doubts about the quality (e.g. when the description text is too short). An icon pair shows that quality is OK for presentation of the item in the DMG-Lib portal as well as in the OAI-PMH-ESE interface – i.e. it will be shown in Europeana after the next harvesting.

From the software implementation point of view a set of quality checker classes was implemented which is able to calculate if/where items can be shown and what must be done in case that an item cannot be shown. Via the icons the efficiency of the work in WP6 was improved. But the quality checker classes are also used in WP2 for quality checks to decide if an item can be added to the OAI interface or not.

Fast adaptability of functionality
For several calculation methods like quality checking as well as the mapping of DMG-Lib metadata to DC or ESE it is useful to have a possibility to adapt the behavior without restart of the application server (at dmg-lib.org Apache Tomcat is used as application server), because that can be done only during nightly maintenance sessions. Therefore a new concept was introduced that allows dynamic loading of classes which contain quality checking or metadata mapping code. This allows fast reactions on problems in the mapping. During the work with the Europeana Content Checker the use of dynamically loadable classes shortened the development cycles.

Caching of Responses
Before the harvesting tests from the ingestion team the calculation of the responses was not cached. Due to the considerable increase of items the calculation time for a complete record list was in May ten and more minutes. Because of the disability to harvest from ingestion team it was assumed that a timeout could be the reason. Therefore a response cache was introduced which can be filled (e.g. at night) and which is then ready to send responses to the Europeana harvester. Using the cache the transfer of the 29MB of data was done in 20 to 40 seconds.

Debug code
Because of the unclear reason of the disability to harvest at different places debug code was integrated into the source code of the OAI-interface. This helped to check/understand the harvesting process performed by Europeana as far as it concerns the DMG-Lib server.
Because the Content Checker did not accept an XML-file which had all the OAI-PMH specific wrapper nodes a stripped version of the XML-files was used for the tests with the content checker. For a more comfortable generation of XML-files to be tested in the content checker some special OAI-PMH response generation switches were introduced. Such stripped XML-files (containing only the ESE-records) could also be sent to the Europeana harvester or to the ingestion team for ingestion by hand.

**Figure 5. Control interface in the ProDB for activating debug code, stripping or caching**

![Control interface](image)

**XML-validation**

We recommend jEdit (an open source editor, see www.jedit.org) with the XML-plugin for checking the validity of your XML-code.

4. **Results**

Currently (Nov 2012), 14,283 DMG-Lib items are visible in the Europeana portal.

An updated dataset of approximately 29,000 DMG-Lib was successfully ingested in November 2012 and will be part of the next publication, due the first week of December 2012.

The monthly harvesting of the continuously growing DMG-Lib database now can be performed automatically by Europeana.
ANNEX I – Well-structured XML-data for automatic harvesting of items by Europeana

<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2012-11-18T22:37:02Z</responseDate>
  <request verb="ListRecords" metadataPrefix="ese">http://www.dmg-
  lib.org/dmglib/OAI-2.0-Server</request>
  <ListRecords>
    <oai:record xmlns:oai="http://www.openarchives.org/OAI/2.0/">
      <header>
        <identifier>oai:dmglib.org:2022</identifier>
        <datestamp>2006-01-12</datestamp>
        <setSpec>doc-type:manim</setSpec>
        <setSpec>doc-type:iactm</setSpec>
      </header>
      <oai:metadata>
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/
xsi:schemaLocation="http://www.europeana.eu/schemas/ese/
http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd">
          <dc:title>Lenkergeradführung</dc:title>
          <dc:creator>Michael Reeßing</dc:creator>
          <dc:date>2006-01-12</dc:date>
          <dc:type>animated image sequence</dc:type>
          <dc:identifier>dmg:2022</dc:identifier>
          <dc:identifier>http://www.dmg-
          lib.org/dmglib/handler?manim=2022</dc:identifier>
          <dc:relation>Ilmenau University of Technology, Germany</dc:relation>
          <dc:rights>http://www.dmg-
          lib.org/dmglib/main/portal.jsp?mainNaviState=site.oairights</dc:rights>
          <europeana:provider>thinkMOTION</europeana:provider>
          <europeana:type>VIDEO</europeana:type>
          <europeana:dataProvider>Digital Mechanism and Gear Library – www.dmg-
          lib.org</europeana:dataProvider>
          <europeana:isShownAt>http://www.dmg-
          lib.org/dmglib/handler?manim=2022</europeana:isShownAt>
          <europeana:isShownBy>http://www.dmg-
          lib.org/dmglib/main/manimViewer_content.jsp?id=2022&amp;skipSearchBar=1</europeana:isShownBy>
        </europeana:record>
      </oai:metadata>
    </oai:record>
  </ListRecords>
</OAI-PMH>