Disclosing the European Library on common visual historical heritage

D6.2 “EURO-Photo Interoperability Framework and news agency module testing report”

<table>
<thead>
<tr>
<th>Deliverable Form</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Reference No.</strong></td>
</tr>
<tr>
<td><strong>Deliverable No.</strong></td>
</tr>
<tr>
<td><strong>Relevant Workpackage:</strong></td>
</tr>
<tr>
<td><strong>Nature:</strong></td>
</tr>
<tr>
<td><strong>Dissemination Level:</strong></td>
</tr>
<tr>
<td><strong>Document version:</strong></td>
</tr>
<tr>
<td><strong>Date:</strong></td>
</tr>
<tr>
<td><strong>Authors:</strong></td>
</tr>
<tr>
<td><strong>Contributors:</strong></td>
</tr>
<tr>
<td><strong>Document description:</strong></td>
</tr>
</tbody>
</table>
Executive Summary

This document summarizes the final assessment activities of EURO-Photo portal prototype (www.EURO-Photo.eu.com), describes News Agencies local modules implementation & assessment and Europeana harvester implementation and initial upload activities.

Chapter 1: EURO-Photo technical assessment describes activities and outcome of the technical working group set up in Copenhagen Meeting in June (composed by EFE and Scanpix as leading partners, LUSA, MTI, Ansa and Expert System). The aim of this group was to work in close cooperation with Expert System on main “open issues” toward the commercial launch of EURO-Photo portal.

Chapter 2: News Agencies local module implementation & assessment describes workflow and individual transferring modules implemented at each Agency site and its assessment.

Chapter 3: Europeana harvester implementation and initial upload describes the harvesting module, the harvesting process and provide a report on uploaded pictures.

Appendix I: Europeana content provision Form
# Table of Contents

LIST OF FIGURES ...................................................................................................... 5
LIST OF TERMS AND ABBREVIATIONS ..................................................................... 6
INTRODUCTION ........................................................................................................ 7

1 EURO-PHOTO FINAL TECHNICAL ASSESSMENT .............................................. 8

1.1 EURO-Photo WORKGROUP set up and approach ........................................... 8
1.2 EURO-Photo WORKGROUP Evaluation and Testing ....................................... 8

2 NEWS AGENCIES LOCAL MODULE DESCRIPTION & ASSESSMENT .............. 11

2.1 ANSA local module description & assessment .............................................. 11
   2.1.1 TECHNICAL BACKGROUND .................................................................... 11
   2.1.2 FIELDS TO BE FILLED ........................................................................... 12
   2.1.3 EXPORTING TO EURO-PHOTO .............................................................. 13
   2.1.4 SAMPLE XML FILE .............................................................................. 14

2.2 Belga local module description & assessment .............................................. 15
   2.2.1 First phase: IPTC meta data definition and picture scanning .................... 15
   2.2.2 Second phase: batch process for archive pictures ....................................... 17

2.3 DPA local module description & assessment ................................................. 21
   2.3.1 Technical background .............................................................................. 21
   2.3.2 Fields to be filled .................................................................................... 22
   2.3.3 The News Agency module ....................................................................... 24
   2.3.4 Sample XML file .................................................................................... 25

2.4 EFE local module description & assessment .................................................. 28
   2.4.1 Workflow .................................................................................................. 28
   2.4.2 Cataloguing pictures ............................................................................... 29
   2.4.3 Exporting to EURO-Photo ..........................................................33
   2.4.4 Example ................................................................................................. 36

2.5 EPA local module description & assessment ................................................. 40
   2.5.1 Technical background .............................................................................. 40
   2.5.2 Fields to be filled .................................................................................... 41
   2.5.3 The News Agency module ....................................................................... 42

2.6 LUSA local module description & assessment .............................................. 43
   2.6.1 Technical background summary of LUSA .............................................. 43
   2.6.2 Handling photos and metadata filling for EURO-Photo ............................. 43
   2.6.3 Integration with EURO-Photo ................................................................. 43

2.7 MTI local module description & assessment ................................................. 45
   2.7.1 Technical background .............................................................................. 45
   2.7.2 Fields to be filled .................................................................................... 45
   2.7.3 The News Agency module ....................................................................... 46
   2.7.4 Sample XML file .................................................................................... 48
2.8  PAP local module description & assessment .......................... 51
  2.8.1  Technical background ............................................................. 51
  2.8.2  Fields to be filled ................................................................. 51
  2.8.3  The News Agency module .................................................. 52

3  EUROPEANA HARVESTING & ON LINE PUBLICATION ................. 54

CONCLUSION .......................................................................................... 61

APPENDIX I : EUROPEANA CONTENT PROVISION FORM ...................... 62
LIST OF FIGURES

Figure 1 : ANSA Storico – screenshot ........................................................................... 11
Figure 2 : PhotoMDE user interface : ............................................................................ 16
Figure 3 : Belga - Use case schema ............................................................................. 17
Figure 4 : Belga Batch process activity schema: ......................................................... 20
Figure 5 : EFE workflow ............................................................................................. 28
Figure 6 : EFE cataloguing pictures ............................................................................ 29
Figure 7 : EFE User Interface ..................................................................................... 30
Figure 8 : EFE Search results ..................................................................................... 30
Figure 9 : EFE - Origin and subject of the picture ....................................................... 31
Figure 10: Location and storage of the original picture .......................................... 32
Figure 11: Data describing the picture itself and any other relevant information .......... 32
Figure 12: MTI Workflow .......................................................................................... 47
Figure 13: PAP - Historians at work............................................................................ 52
Figure 14: PAP Screen shoots .................................................................................... 53
Figure 15: Aggregators in the Europeana organisational model ............................... 54
Figure 16: EURO-Photo pictures shown at Europeana .............................................. 57
Figure 17: EURO-Photo picture shown at Europeana ............................................... 58
Figure 18: Europeana links EURO-photo (ShownAt mode) ...................................... 59
LIST OF TERMS AND ABBREVIATIONS

EDM  Europeana Data Model (EDM)
ESE  Europeana Semantic Elements (ESE)
OAI-ORE  Open Archives Initiative - Object Reuse and Exchange (OAI-ORE) - defines standards for the description and exchange of aggregations of Web resources.
OAI-PMH  Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). is a low-barrier mechanism for repository interoperability. *Data Providers* are repositories that expose structured metadata via OAI-PMH. *Service Providers* then make OAI-PMH service requests to harvest that metadata. OAI-PMH is a set of six verbs or services that are invoked within HTTP.
INTRODUCTION

This document summarizes the final assessment activities of EURO-Photo portal prototype (www.EURO-Photo.eu.com), describes News Agencies local modules implementation & assessment and Europeana harvester implementation and initial upload activities.

Chapter 1: EURO-Photo technical assessment describes activities and outcome of the technical working group set up in Copenhagen Meeting in June (composed by EFE and Scanpix as leading partners, LUSA, MTI, Ansa and Expert System). The aim of this group was to work in close cooperation with Expert System on main “open issues” toward the commercial launch of EURO-Photo portal.

Chapter 2: News Agencies local module implementation & assessment describes workflow and individual transferring modules implemented at each Agency site and its assessment.

Chapter 3: Europeana harvester implementation and initial upload describes the harvesting module, the harvesting process and provide a report on uploaded pictures.
1 EURO-Photo Final technical assessment

1.1 EURO-Photo WORKGROUP set up and approach

The workgroup was composed by a subset of EURO-Photo participating agencies namely:

- EFE & Scanpix as team leaders
- LUSA,
- MTI,
- ANSA

The aim of his working group was to make a punctual review of each functionality of EURO-Photo portal and an overall assessment work in close cooperation with Expert System in order to sort out “open issues” toward the commercial launch of EURO-Photo portal.

1.2 EURO-Photo WORKGROUP Evaluation and Testing

Activity started from EURO-Photo rel. 3.0 in order to assess the system toward the final release (4.0)

In release 3.1 the following topics were addressed:

- Search page was split up into two distinct pages (simple search/ semantic search) proposing a traditional solution (simple search) without semantics as default for occasional user and letting advanced user benefitting of the more powerful semantic approach. The semantic search page was enriched with a quick guide explaining what is "semantic search", how to use it, when it's better using this search than the "simple" search.
All metadata fields were included in “simple search”: City, Country, Keywords ... not only Title and Description, to make search activities more straight and intuitive (achieving in any case good results without using more powerful features).

GUID unique identifier become also searchable directly from “simple search” from search field.

Graphic layout minor improvements (e.g. logos resizing and alignment, icons resizing/replacement, font changes, field repositioning and resizing, etc.).

Paging in pictures visualization was made more intuitive and efficient (either grid view or list view).

User registration procedure was revised and completed (e.g. “lost password recoupment procedure” was missing).

Picture uploading process from Agencies was improved sending a more comprehensive report about the uploading process.

In release 3.2 the following topics were addressed:

- Enhancements and rationalization of query construction, integration of picking-up of fields from faceting or picture metadata with manual entering.

- Enhancements of “simple search” to fully support Boolean operators and their nesting with brackets, in order to provide to “traditional users” a more familiar way to make advanced search instead of using the more powerful “semantic search”.

- Overall search response time was dramatically reduced optimizing also pictures visualization to reduce delays.
Graphic optimization of the interface to an agreed reference resolution of 1280 X 1024x1024.

Help on-line revision and extension.

In release 4.0 the following topics were addressed:

- Automatic IPTC categorization process was revised and tuned in close cooperation with News Agencies, taking into account the fact that descriptions of historical pictures is often very short and Google automatic translation in rather imprecise. In any case we decided to limit automatic generation to the second level of IPTC taxonomy.

- We improved automating Google translation of Title and Description implementing a pre-processor to handle proper nouns.

- The above mentioned pre-processor is using either the single picture metadata (e.g. City, Person_Shown) or proper nouns lists provided from Agencies. In order to avoid undesirable translation (e.g. Person name or small city name) or to force, when appropriate, predefined translation (e.g major cities)

- Starting slide show has been replaced with a new video clip provided by Ansa.
2 News Agencies local module description & assessment

2.1 ANSA local module description & assessment

2.1.1 TECHNICAL BACKGROUND

The ANSA news agency is using ANSASTORICO web application for managing, archiving and serving photos. ANSASTORICO is used by photo editors and people who work for the ANSA photo archives for adding photos and metadata to photos. It is based on Open Source technology like the web server Apache, MySql, PHP5, Zend Framework and GD Library. The photos are stored in their original size in JPEG format, all the text is stored embedded IPTC. Along with the original size are stored a medium (about 1024x768 pixels) and small (about 96x96 pixels) resolution also. Another features is search functionality. The research is based on the photo’s metadata. For example, users can search by a category, subcategory, agency, place, publishing date, etc.

Figure 1: ANSA Storico – screenshot
As you can see in the figure above ANSASTORICO shows the search results as a list of thumbnails. Under each thumbnail are displayed some information about the photo as title, description, event date, etc. By clicking it is possible to see the photo medium resolution and all others information.

The automatic categorization system is based on Mysql database.

All the metadata is stored in the JPEG file and ANSASTORICO is able to use it correctly.

Of course it is possible to export these informations to serve other applications or to create a sidecar XML format with metadata.

2.1.2 FIELDS TO BE FILLED

During the cataloguing, ANSASTORICO try to fill in as many fields as we can.
The following figure shows the data for database photo table structure.

<table>
<thead>
<tr>
<th>Campo</th>
<th>Tipo</th>
<th>Collation</th>
<th>Attributi</th>
<th>Null</th>
<th>Predefiniti</th>
<th>Extra</th>
</tr>
</thead>
<tbody>
<tr>
<td>id_foto</td>
<td>int(11)</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td>AUTO_INCREMENT</td>
</tr>
<tr>
<td>uuid</td>
<td>char(36)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>id_destinazione</td>
<td>int(11)</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>titolo</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>didascalia</td>
<td>text</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>titolo_en</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>didascalia_en</td>
<td>text</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lungo_scritto</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>antropomorfo</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eticheta</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tipologia_autore</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>istruzioni_speciali</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stato_lavorazione</td>
<td>enum('0','1','2','3','4','5','6','7','8','9')</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nazionalità</td>
<td>char(1)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>data</td>
<td>datetime</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>categoria</td>
<td>varchar(128)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>estetica</td>
<td>varchar(8)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>id_agenzia</td>
<td>int(4)</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>keywords</td>
<td>varchar(256)</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>width</td>
<td>int(4)</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>height</td>
<td>int(4)</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>partner</td>
<td>enum('0','1','2')</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lar</td>
<td>enum('0','9')</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>id_utente</td>
<td>int(11)</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>id_utente_ultima_modifica</td>
<td>int(11)</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>data_inserimento</td>
<td>datetime</td>
<td></td>
<td></td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>data_ultima_modifica</td>
<td>timestamp</td>
<td>on update CURRENT_TIMESTAMP</td>
<td>No</td>
<td>0000.00.00 00:00:00.00</td>
<td>on update CURRENT_TIMESTAMP</td>
<td></td>
</tr>
<tr>
<td>keywords2</td>
<td>text</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>partner1</td>
<td>enum('0','1','2')</td>
<td>latin1_swedish_ci</td>
<td>No</td>
<td>None</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.1.3 EXPORTING TO EURO-PHOTO

Every picture in the ANSASTORICO repository is available as the object itself and an XML containing all the metadata for the photo. You can select single photos, a group of photos and a photos set that are the result of a search.

The export result is a zip file containing a set of file pairs: jpeg file and relative xml file. Both file name
2.1.4 SAMPLE XML FILE

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<ansastorico>
  <photo>
    <guid>tag:ansa.it,2011-08-03:69d96d606d1211e1ae16bc305bd461d9</guid>
    <id>73648</id>
    <bustinaId>11778</bustinaId>
    <title>Maria Callas abbandona il Teatro dell'Opera interrompendo la "Norma"</title>
    <description>Il pubblico all'esterno del Teatro dell'Opera dopo l'interruzione de la "Norma" da parte del soprano Maria Callas, Roma 2 gennaio 1958. ANSA</description>
    <titleEn>Maria Callas leaves the Teatro dell'Opera interrupting "Norma"</titleEn>
    <descriptionEn>The public outside the Teatro dell'Opera after soprano Maria Callas interrupted "Norma", Roma 2 January 1958. ANSA</descriptionEn>
    <location>RM</location>
    <antroponimi />
    <enti />
    <authorTipology />
    <specialInstructions />
    <workingStatus>7</workingStatus>
    <country>ITA</country>
    <city>Roma</city>
    <date>1958-01-02 00:00:00</date>
    <category>ACE</category>
    <subcategory>01017000</subcategory>
    <keywords>cronaca, esterno, orizzontale news, exterior, horizontal</keywords>
    <width>3571</width>
  </photo>
</ansastorico>
```
2.2 Belga local module description & assessment

(Authors: Vincent Fréry, Stefaan Melis, Wuytack Tom)

This section deals with the technical aspect of the EURO-PHOTO implementation of local modules at Belga, NOT the selection phase of the images which comes first in the Belga workflow. Neither does this document enter into detail on the general tools/platform in use by Belga for its picture production and sales. This description has been included in an earlier deliverable. The focus is thus on the local modules developed to support interoperability. These will be explained following the picture archive process. This picture archive process is done in 2 phases.

2.2.1 First phase: IPTC meta data definition and picture scanning

2.2.1.1 Define IPTC metadata

IPTC metadata are defined for each archive picture. Once the metadata defined, they are stored into a database. In order to do this, in collaboration with our end users, a web
client application called PhotoMDE (Meta Data Editor) was developed specifically for the archiving project. PhotoMDE is a standalone Java based web client that offers a user friendly way to edit, create, delete and update IPTC meta data.

*PhotoMDE user interface*:

![PhotoMDE User Interface](image)

**Figure 1** : Belga PhotoMDE user interface :

### 2.2.1.2 Archive picture scanning

Once archive picture metadata are stored, each picture is scanned and stored on a central repository. In order to link the scanned picture to the stored previously IPTC information into the database, the name of each file matches the corresponding IPTC record id.
2.2.2 Second phase: batch process for archive pictures.

Once an image set is scanned and his relative metadata inserted into a database, two batch processes wrote in per language automate the remaining work.

This is a snipped code from the first perl script:

```perl
# Make the conversion with ImageMagik.
sub convert_pix {
    my ( $source, $dest, $conversion ) = @_;    
    my $command = "\"C:/Program Files (x86)/ImageMagick-6.6.9-Q16/convert\" \"$source\" -resize $CONVERSION $dest";
    open( CMD, "$command |" );
    close(CMD);
    log_message("$source converted into $dest.");
}
```

# Insert meta data for all pictures
sub insert_meta {
  my ($pix_dir) = @_;
  print("Insert meta for $pix_dir directory...");
  foreach $f ( get_pictures_file_name($pix_dir) ) {
    insert_meta_into_picture( $pix_dir . "/" . $f, get_picture_archive_IPCT( trim_zero_picture_id($f) ) ) unless
    (( $f eq "." ) || ( $f eq "..");
    print("[OK]\n");
  }
}

sub get_IPTC_date {
  my ($date) = @_;
  @d = split('/', $date);
}

A first batch process will insert the IPTC metadata from the archive picture database into all picture headers.

All images are then copied in a lower resolution format (600x600) into a folder to be sent to Euro-photo.

Then, the original image set is covered by a process that checks image headers and store that information into the Belga image database.

Here is an example of picture information that can be retrieved through de database:

<table>
<thead>
<tr>
<th>COLUMN_NAME</th>
<th>DATA_TYPE</th>
<th>NULLABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMAGE_ID</td>
<td>NUMBER(9,0)</td>
<td>No</td>
</tr>
<tr>
<td>USER_ID</td>
<td>CHAR(6 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>FDISK_NO</td>
<td>NUMBER(9,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>PDISK_NO</td>
<td>NUMBER(9,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>TDISK_NO</td>
<td>NUMBER(9,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>FULLNAME</td>
<td>VARCHAR2(128 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>PREVIEWNAME</td>
<td>VARCHAR2(128 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>THUMBNAME</td>
<td>VARCHAR2(128 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>PROTECTED</td>
<td>CHAR(1 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>ARCHIVE</td>
<td>CHAR(1 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>WIDTH</td>
<td>NUMBER(9,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>HEIGHT</td>
<td>NUMBER(9,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>RDATE</td>
<td>DATE</td>
<td>Yes</td>
</tr>
<tr>
<td>DATABASE</td>
<td>NUMBER(6,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>-----</td>
</tr>
<tr>
<td>NAME</td>
<td>VARCHAR2(256 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>CAPTION</td>
<td>VARCHAR2(2000 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>AUTHOR</td>
<td>VARCHAR2(64 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>CATEGORY</td>
<td>VARCHAR2(64 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>CATEGORY_ID</td>
<td>NUMBER(38,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>SUPCATEGORY</td>
<td>VARCHAR2(64 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>CITY</td>
<td>VARCHAR2(128 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>COUNTRY</td>
<td>VARCHAR2(128 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>CREDIT</td>
<td>VARCHAR2(64 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>CREDIT_ID</td>
<td>NUMBER(38,0)</td>
<td>Yes</td>
</tr>
<tr>
<td>SOURCE</td>
<td>VARCHAR2(64 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>UPDATE_DATE</td>
<td>DATE</td>
<td>Yes</td>
</tr>
<tr>
<td>SEARCH_TEXT</td>
<td>VARCHAR2(4000 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>KEYWORDS</td>
<td>VARCHAR2(1000 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>THEME</td>
<td>VARCHAR2(128 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>STYLE</td>
<td>CHAR(1 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>REFERENCE</td>
<td>VARCHAR2(512 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>EXPORT_DESTINATIONS</td>
<td>VARCHAR2(512 BYTE)</td>
<td>Yes</td>
</tr>
<tr>
<td>MEDIA_ID</td>
<td>NUMBER</td>
<td>Yes</td>
</tr>
</tbody>
</table>

A second batch process retrieve the original image id and insert a copy into the headers of images to be send to EURO-Photo.

Finally, the low resolution image set that contains the reference to the original image is sent to EURO-Photo via FTP.

Batch process activity schema:
Figure 3: Belga Batch process activity schema:
2.3 DPA local module description & assessment

2.3.1 Technical background

The German Press Agency dpa is operating with several software products for production, archiving and selling/offering images.

Newsroom-editors work with an own system that was developed by dpa, called “ines” (integrated news editing system”).

The archive production is splitted to off-line working and on-line working.

Off-line: Documentation specialists of the image archive screen and select prints and slides/negatives and hand them over to the dpa Photoservice for scanning high-resolution files. These files are being edited off-line with FotoStationPro in IPTC-IIM (Record 2) standard.

On-line: The Photoservice imports the files into the platform (database) without any further information. The documentation specialists are doing the keywording online via FAT-Client or web browser.

dpa stores all images in an asset management system, based on oracle database and Linux operations system on a server farm with enterprise hardware/mass storage system. Three webservers offer the content in the internet for online research by clients.

All files are stored in original size, compressed as JPEG and with no ICC color profile, in RGB. After import to the database the IPTC meta data are separated in the database for additional editing and corrections. A bidirectional translation (DE <-> EN) is only used for keywords. On download the current database data will be embedded back into the download file. At same time all EXIF and XMP datas will be deleted from the file header.

Mostly all data are full text searchable. There is no hierarchy and tree or catalogue for keywords and search items. Search is oracle technology.

Dpa is using also extended fields for sales and trade restriction information. This is necessary for the international client rights management.

For mass export we have the opportunity to deliver a separated XML file for each image (e.g. export to EURO-PHOTO).
2.3.2 Fields to be filled

dpa is using the following fields:

IPTC-IIM (Record 2)
2:105 Headline (german)
2:106 Headline (englisch, internal usage)
2:120 Caption (german)
2:220 Caption (englisch, internal usage)
2:020 Suppl. Category (optional)
2:025 Keywords (german)
2:203 Keywords (englisch, internal usage)
2:080 Byline
2:055 Date created
2:090 City
2:095 Province
2:100 Country code (optional)
2:101 Country
2:200 Continent (internal usage)
2:015 Category (optional)
2:040 Special instructions (optional)
2:122 Caption writer (optional)
2:110 Credit
2:115 Source
2:116 Copyright notice
2:131 Image orientation (optional)
Additional fields in own standard

Imaged person

Concept (for creative content)

Type of picture (e.g. Portrait, Group, Night scene)

Emotion/Association (for creative content)

Color

Photo series

ID

SupplierID

ExternalID

Date stored

Image size width

Image size height

Color depth

Color space

File size

Image size

Original/draft (print, slide)

Crop in percent

Folder

Trade right

Memo

Priority

Countrygroup

Restrictions

XMP schema: dpa is not working with XMP.
2.3.3 The News Agency module

The News agency module has been developed by the agency according to integration software specification in D3.5 document.

After the human selection for EURO-PHOTO, all selected images will get an additional keyword “EURO-Photodb” for identification purposes. No more work is needed, because all translation and keyword work is already done.

IT service is exporting the preview version, incl. XML file. Each file pair will be transmitted via ftp to Experts System’s ftp server. (00000.jpg & 00000.xml)
2.3.4 Sample XML file

```xml
<?xml version="1.0" encoding="UTF-8"?>
<picture_to_insert mediatype="image" partnerId="22004">
<originalId><![CDATA[urn:newsml:picture-alliance.com:20030101:0000452745]]></originalId>
<technicalData width="6032" height="4392" unit="pixel" cDepth="8" cSpace="Grayscale" color="4C" sHires="false">
  <mimeType><![CDATA[image/jpeg]]></mimeType>
  <url protocol="file" path="23835514.jpg"/>
</technicalData>
<formalData>
  <eaw><![CDATA[ba]]></eaw>
  <urgency><![CDATA[1]]></urgency>
  <npi_report_id><![CDATA[0]]></npi_report_id>
</formalData>
<contentData>
  <title lang="de"><![CDATA[Wahlplakate für Wahl zur Verfassungsgebenden Landesversammlung des neuen Südweststaats]]></title>
  <title lang="en"><![CDATA[Election posters for the election of the constituent assembly of the new south-western state]]></title>
  <caption lang="en"><![CDATA[Election posters for the election of the "constituent assembly" of the new south-western state are hung up on a board in Stuttgart on 9 March 1952. In a referendum on 9 December 1951, the majority of the population of the three south-western states of Germany declared for the formation of a common federal state. In the end, a federal state, called Baden-Württemberg, consisting of the federal states Baden, Württemberg-Baden and Württemberg-Hohenzollern, was found on 25 April 1952.]]></caption>
  <keyword lang="de"><![CDATA[Bretterwand]]></keyword>
  <keyword lang="de"><![CDATA[kleben]]></keyword>
  <keyword lang="de"><![CDATA[Landesregierung]]></keyword>
  <keyword lang="de"><![CDATA[Wahlkampf]]></keyword>
  <keyword lang="de"><![CDATA[Politiker]]></keyword>
  <keyword lang="de"><![CDATA[Politik]]></keyword>
  <keyword lang="de"><![CDATA[Wahlen]]></keyword>
  <keyword lang="de"><![CDATA[historisch]]></keyword>
  <keyword lang="de"><![CDATA[Wand]]></keyword>
  <keyword lang="de"><![CDATA[Wählen]]></keyword>
  <keyword lang="de"><![CDATA[Geschichte]]></keyword>
  <keyword lang="de"><![CDATA[Gruppe]]></keyword>
  <keyword lang="de"><![CDATA[Personen]]></keyword>
  <keyword lang="de"><![CDATA[EURO-Photodb]]></keyword>
  <keyword lang="de"><![CDATA[Bundesland]]></keyword>
</contentData>
</picture_to_insert>
```
2.4 EFE local module description & assessment

2.4.1 Workflow

The workflow used by Agencia EFE for the EURO-Photo project is graphically described in the next image and should be kept in mind to understand the software used and specifically developed as the “News Agency Module” for EFE.

![Figure 4: EFE workflow](image)

There are two main software frameworks:

- The interactive software used for entering metadata describing pictures. It’s the software used by people in the Archive Department.

- The automatic processes developed for this project that will be able to take all the pictures fully captioned and keyworded and put it in the agreed exchange format for entering the EURO-Photo portal.
2.4.2 Cataloguing pictures

Agencia EFE is using proprietary software to add textual metadata into the digitalised and selected pictures involved in the project. This software is called Fotomind and was developed years ago by the Spanish company Gesfor. The software used by the people in Archive Documentation Department provides a way of browsing, selecting and tagging every picture with different kind of metadata in a well-structured manner. The backend solution is based on an Oracle database environment managing all the information with clients (people in the Archive Documenting department) accessing to this server components.

![Diagram](image)

**Figure 5: EFE cataloguing pictures**

This software provides an environment where pictures can be searched in order to be captioned and keyworded. The first step in the workflow is locating the pictures recently scanned, which are entered into the file repository and into de database properly marked. The database may be queried through a search tool which provides all kind of operators and fields for searching, as seen in the figure below. Conditions may be added for convenience and used in the most precise manner to get the best result for the search.
As result of the search, you get a window split in two main parts. On the left, you’ll find a browser for the pictures matching the query and on the right you can see detailed view for the picture selected. Here you can decide to “Edit” that information to enrich textual content of the photo.
When pictures are located, they are fully documented using all the information available that suites every field in the database model. These fields are grouped in three different tabs:

Data about the origin and subject of the picture (tab “Origen/Clasificacion”). There you can add information about place and date of the picture, photographer, source and credit and main subject describing the photo.

![Figure 8: EFE - Origin and subject of the picture](image-url)

Data about location and storage of the original picture, if it’s not a digital picture (tab “Características”). Here you can also add information about entities, companies and people related to the picture, if any.
Figure 9: Location and storage of the original picture

Data describing the picture itself and any other relevant information (tab “Pie/Descripción”). Here you’ll find the caption of the photo and more, like extra categorization, keywords, and other relevant descriptive information that applies.

Figure 10: Data describing the picture itself and any other relevant information

For the EURO-Photo project, only caption is manually entered in Spanish and English. All other translations are done using controlled vocabulary in both languages, as explained later.
2.4.3 Exporting to EURO-Photo

Every picture in the Archive repository is available as the object itself and an XML (proprietary format) containing all the metadata for the photo. We’ll use that XML as the starting point to get to the agreed metadata file format, also in XML.

Textual information added to the pictures needs to be reformatted in order to agree the exchange format defined by the members of the project. The fields defined for this format are shown in the table below, where you can find when EFE is providing such information and if it’s filled by human people or automatically added based on post-processing actions explained later.
<table>
<thead>
<tr>
<th>Field name</th>
<th>Mandatory</th>
<th>Included</th>
<th>How</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language independent</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Image GUID</td>
<td>M</td>
<td>Y</td>
<td>Internal</td>
</tr>
<tr>
<td>IPTC Subject Code</td>
<td>SR</td>
<td>Y</td>
<td>Post-process</td>
</tr>
<tr>
<td>DateCreated</td>
<td>SR</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Creator</td>
<td>R</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Source</td>
<td>M</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td><strong>National Language</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>M</td>
<td>Y</td>
<td>Fixed</td>
</tr>
<tr>
<td>Headline</td>
<td>M</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Caption/Description</td>
<td>M</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Keywords</td>
<td>SR</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>City</td>
<td>SR</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>State/Province</td>
<td>R</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Location Shown</td>
<td>R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person Shown in the Image</td>
<td>R</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Name of Organisation Featured in the Image</td>
<td>R</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Instructions</td>
<td>R</td>
<td>Y</td>
<td>Fixed</td>
</tr>
<tr>
<td>Creator’s Jobtitle</td>
<td>R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copyright Notice</td>
<td>M</td>
<td>Y</td>
<td>Fixed</td>
</tr>
<tr>
<td><strong>English Language</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>M</td>
<td>Y</td>
<td>Fixed</td>
</tr>
<tr>
<td>Headline</td>
<td>M</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Caption/Description</td>
<td>M</td>
<td>Y</td>
<td>Manual</td>
</tr>
<tr>
<td>Keywords</td>
<td>SR</td>
<td>Y</td>
<td>Post-process</td>
</tr>
<tr>
<td>City</td>
<td>SR</td>
<td>Y</td>
<td>Post-process</td>
</tr>
<tr>
<td>State/Province</td>
<td>R</td>
<td>Y</td>
<td>Post-process</td>
</tr>
<tr>
<td>Location Shown</td>
<td>R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person Shown in the Image</td>
<td>R</td>
<td>Y</td>
<td>Post-process</td>
</tr>
<tr>
<td>Name of Organisation Featured in the Image</td>
<td>R</td>
<td>Y</td>
<td>Post-process</td>
</tr>
<tr>
<td>Instructions</td>
<td>R</td>
<td>Y</td>
<td>Fixed</td>
</tr>
<tr>
<td>Creator’s Jobtitle</td>
<td>R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copyright Notice</td>
<td>M</td>
<td>Y</td>
<td>Fixed</td>
</tr>
</tbody>
</table>
Last column shows how the field is filled:

- **Internal** means this is a field not editable by human and it’s assigned by the editing framework.

- **Fixed** means its value is fixed for every picture. The content is defined in the internally post-process phase.

- **Manual** means that the content of the field relies on human editing tasks.

- **Post-process** means that the value is filled automatically extracting some “manual” fields that are translated or interpreted in order to get the correct values.

The most important part of this phase of the workflow is understanding how the “post-process” is filled using the information added manually in the editing framework. For those fields not mapped directly to the fields available in the editing software, we write all the relevant information in the wide purpose field “Descripción” (“Description”). In the post-process, that field is parsed in order to extract specific parts of it and give the correct structure to for the EURO-Photo-XML exchange format.

- **IPTC Subject codes**: Description field is parsed looking for the label “CATEGORIAS:” and from there we read textual label for the subject of the picture. As it comes from controlled vocabulary, we can extract the IPTC Subject Code and also the translation of the name of the subject code to English. The textual values of the different levels are also written as keywords.

- **Keywords**: Description field is parsed looking for the label “PALABRAS CLAVE:” taking every word or group of words separated with commas as a single keyword. They will be written as “Keywords” in the Spanish section of the XML. There’s a set of keywords used from a Controlled vocabulary, available in English too. Every Spanish keyword in the picture present in the controlled list will be written (translated version of the list) as “Keywords” in the English section of the XML.

- **Location names**: As the information is manually added with the name of the cities, regions and countries only in Spanish, we use automatic translation via “Linked Data API” of the public site Geonames.org to get the English version of the geographic names and put it into the English section of the exported file.

- **Persons shown**: for the English section of the XML we remove all words written in Spanish leaving just the name itself.
Proprietary XML file with metadata:

```xml
<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
<Data>
  <Record>
  <cod_doc>4729566</cod_doc>
  <revisado>20110517</revisado>
  <pais_cod>BEL</pais_cod>
  <ciudad>BRUSELAS</ciudad>
  <lugar>0</lugar>
  <autor />
  <agencia>EFE/EFE</agencia>
  <ftrans>19771116</ftrans>
  <fcreat>19771116</fcreat>
  <orient>0</orient>
  <color>3</color>
  <tamano>2734x3543</tamano>
  <plano>Plano General</plano>
  <pie>VIAJE OFICIAL DE LOS REYES A BÉLGICA: Bruselas 16-11-1977.- Los reyes Juan Carlos y Sofía, acompañados de los Soberanos belgas, Fabiola y Baudouin, saludan desde el balcón del Ayuntamiento al público congregado para darles la bienvenida a la ciudad. EFE/yv BELGIUM-SPAIN-SPANISH KING AND QUEEN VISITS: BRUSSELS (BELGIUM), 16/11/1977.- Spanish King Juan Carlos I and Queen Sofia are seen together with King of the Belgians Baudouin and Queen Fabiola, at the balcony of the Brussels Town Hall during their welcome ceremony. EFE/fs</pie>
  <tipo>Biográfico</tipo>
  <siglas_ent />
  <descripcion>CATEGORÍAS: Política: Asuntos exteriores PALABRAS CLAVE: Transición democrática, realeza, gesto: saludando. casa real española, relaciones bilaterales países, viaje oficial Bélgica yv EUROPHOTO PROJECT</descripcion>
  <clas_cod>11002000000000</clas_cod>
  <identificacion>ESCANEADA DIAPOSITIVA</identificacion>
  <estado>1</estado>
  <localizacion>P.- 2065/78.1</localizacion>
  <ubicacion />
  <copyright>1</copyright>
  <ambito>1</ambito>
  <tarifa>0</tarifa>
  <estilo>Vertical</estilo>
  <personas>JUAN CARLOS I DE ESPAÑA, JUAN CARLOS DE BORBON, SOFIA DE ESPAÑA (ESPOSA JUAN CARLOS I)</personas>
  <cod_agencia>70</cod_agencia>
  <agencia_name>EFE</agencia_name>
  <clasificacion>Asuntos externos</clasificacion>
  <desc_ent />
</Record>
</Data>
```
EURO-Photo XML File

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<EURO-Photo_record>
  <header>
    <identifier>EFE4729566</identifier>
    <datestamp>2011-09-03</datestamp>
    <setSpec>EURO-Photo</setSpec>
    <setSpec>europeana</setSpec>
  </header>
  <record xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:photoshop="http://ns.adobe.com/photoshop/1.0/">
    <Iptc4xmpExt:DIGImageGUID>EFE4729566</Iptc4xmpExt:DIGImageGUID>
    <Iptc4xmpExt:SubjectCode/>
    <rdf:Bag>
      <rdf:li>11000000</rdf:li>
      <rdf:li>11002000</rdf:li>
    </rdf:Bag>
  </record>
</EURO-Photo_record>
```
Los reyes Juan Carlos y Sofía, acompañados de los Soberanos belgas, Fabiola y Balduino, saldan desde el balcón del Ayuntamiento al público congregado para darles la bienvenida a la ciudad.

EFE/vv
BELGIUM-SPAIN-SPANISH KING AND QUEEN VISITS

BRUSSELS (BELGIUM), 16/11/1977. - Spanish King Juan Carlos I and Queen Sofia are seen together with King of the Belgians Baudouin and Queen Fabiola, at the balcony of the Brussels Town Hall during their welcome ceremony.

EFE/fs

- DIPLOMACY
- GREETING
- ROYALTY
- POLITICS
- GESTURE
- DEMOCRATIC TRANSITION

BRUSSELS

- JUAN CARLOS I DE ESPAÑA
- JUAN CARLOS DE BORBON
- SOFIA DE ESPAÑA

Editorial use only. For any other use an additional licensing is required. Contact Agencia EFE for further assistance.

(c) Agencia EFE. All rights reserved.
2.5 EPA local module description & assessment

2.5.1 Technical background

The European pressphoto agency (epa) is using Fotoware products as Editorial Workflow System for EURO-Photo, (managing, archiving and serving photos) and Adobe Photoshop CS3 as Picture editing Software.

The Software FotoStation7 (photo processing & metadata filing software) is used by our editors for adding/changing metadata to photos.

Further epa use following Fotoware software: Colorfactory (Photo enhancement and workflow automation software), Index Manager (used for indexing, and searching), and Distribution Manager (used for distributing photos to our clients).

The photos are stored in their original size in JPEG format, all metadata is stored embedded IPTC-IIM and XMP.

Our editorial system has a database of Countries, Country Code, Categories /Supplemental Categories and Scene Codes.
### 2.5.2 Fields to be filled

EPA provides the necessary metadata inside the jpg. (see table below) included at minimum, the mandatory fields.

<table>
<thead>
<tr>
<th>IPTC Label (1)</th>
<th>Identifier (1)</th>
<th>IPTC IIMv4 field name (2)</th>
<th>Size</th>
<th>Field</th>
<th>(3)</th>
<th>EPA delivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language independent</td>
<td>Digital Image GUID</td>
<td>Iptc4xmpExt:DigImageGUID</td>
<td>32</td>
<td>2:118</td>
<td>M</td>
<td>YES - epa Imagenumber (epaXXXXXXXX) in Field 2:118 (Userdefined118)</td>
</tr>
<tr>
<td></td>
<td>IPTC Subject Code</td>
<td>Iptc4xmpCore:SubjectCode</td>
<td>Subject Reference</td>
<td>8</td>
<td>2:012</td>
<td>SR</td>
</tr>
<tr>
<td>DateCreated</td>
<td>photoshop:DateCreated</td>
<td>Date Created</td>
<td>8</td>
<td>2:055</td>
<td>SR</td>
<td>YES</td>
</tr>
<tr>
<td>Creator</td>
<td>dc:creator</td>
<td>By-line (Author)</td>
<td>32</td>
<td>2:080</td>
<td>R</td>
<td>YES</td>
</tr>
<tr>
<td>Country Code</td>
<td>Iptc4xmpCore:CountryCode</td>
<td>Country code</td>
<td>3</td>
<td>2:100</td>
<td>SR</td>
<td>YES</td>
</tr>
<tr>
<td>Source</td>
<td>photoshop:Source</td>
<td>Source</td>
<td>32</td>
<td>2:115</td>
<td>M</td>
<td>YES</td>
</tr>
</tbody>
</table>

#### English Language

| Language | dc:language | Language identifier | 3    | 2:135 | M   | YES (only english) |
| Headline | photoshop:Headline | Headline | 256  | 2:105 | M   | YES |
| Caption/Description | dc:description | Caption/Abstract | 2000 | 2:120 | M   | YES |
| Keywords | dc:subject | Keywords | 64   | 2:025 | SR  | YES |
| City | photoshop:City | City | 32   | 2:090 | SR  | YES |
| State/Province | photoshop:State | Province/State | 32   | 2:095 | R   | YES |
| Location Shown | Iptc4xmpExt:LocationShown |  | 256  | R     | NO |
| Person Shown in the Image | Iptc4xmpExt:PersonInImage |  | 256  | R     | YES |
| Name of Organisation Featured in the Image | Iptc4xmpExt:OrganisationInImageName |  | 256  | R     | NO |
| Instructions | photoshop:Instructions | Special Instructions | 256  | 2:040 | R   | YES |
| Creator’s Jobtitle | photoshop:AuthorsPosition | By-line Title(Author position) | 32   | 2:085 | R   | YES |
| Copyright Notice | dc:rights | Copyright Notice | 128  | 2:116 | M   | YES |

(3) M = Mandatory - SR = Strongly Recommended - R = Recommended

**XMP schema:** epa is mapping all IPTC fields to the designated XMP fields.
2.5.3 The News Agency module

The News agency module has been developed by agency according to integration software specification in D3.5 document.

After the human selection for EURO-PHOTO, all selected images will be scanned, photoshoped (only slight cropping...) and get their Metadata and also an additional keyword “historical” for identification purposes.

IT service is exporting the preview version, incl. XML file. Each file pair will be transmitted via ftp to Experts System’s ftp server. During this phase the picture is resized. The JPEG size is defined as 600 pixels on the longer side.

The file name is set to its original epa specific filename with “epa” prefix.

Sample XML file: epa is not working with XML Files. All Metadata is stored embedded IPTC-IIM and XMP.
2.6  LUSA  local module description  & assessment

2.6.1  Technical background summary of LUSA

Lusa is using News Asset Agency Edition, a commercial product of the software company ATC. This software is able to receive, manage, archive and distribute several types of content items like: text, photo, video, etc.

All content items types share the same metadata fields and all tabulated metadata like countries, keywords, subjects, etc exists on the system database on Portuguese and English. Any journalist, editor, archivist is able to import content items to the system and handle it through the system. While the content has not been archived the content can be changed and metadata is filled.

Our photo archivists use the News Asset client application to import photos to the system. The photos are edited with Photoshop that is integrated with the main application. On the system the photos are stored in their original format in JPEG format, but all metadata is resident on the database.

Only an editor approved content item is archived on the system and able to be distributed by several channels that are implemented.

2.6.2  Handling photos and metadata filling for EURO-Photo

After the treatment of the image in Photoshop, respecting the minimum size, the metadata is introduced into the maximum possible fields, with the introduction of mandatory keyword "EURO-Photo" for the system to recognize that the file is from EURO-Photo service, although this keyword is no longer exported along with other metadata. The metadata introduced also respects a criteria for the future search. Example: more than 70% of the Lusa pictures have origin in Lisbon, so to see pictures about Lisbon it’s only necessary to introduce the keyword “City living”.

2.6.3  Integration with EURO-Photo

There have been some developments on the system to provide an alternative language for each content item. This was required for EURO-Photo in order to provide the native (Portuguese) and alternative (English) language for each digitalized photo.

LUSA uses mainly the NewsML1.2 format specification, so it has been developed a XSLT to transform the NewsML1.2 fields to the format specified in D3.5 document.
Using the main feature of the system that all tabulated metadata exists on both languages: Portuguese and English, and with minor adjustments on the export process, LUSA is able to provide the required fields on both languages like specified in D3.5.

When a bulk of photos is ready to be sent to EURO-Photo, the photos are exported, from the system, to a folder, using the XSL transformation. On the export process all filenames receive the prefix "LUS_". Only the photos that are archived and that belong to the EURO-Photo Service are exported in this process. Two files are generated for each photo on this process, the JPG format file with the 600px resolution on the longer side without watermark and a XML file that complies with the format specified in D3.5 document.

When the export completes the photos are manually sent by ftp using any available ftp client.
2.7 MTI local module description & assessment

2.7.1 Technical background

The Hungarian News Agency is using Fotoware products for managing, archiving and serving photos.

The software FotoStation is used by photo editors and people who work for the Photo Archives for adding metadata to photos.

Further elements of the Fotoware software family are: Index Manager (used for indexing, and searching), Fotoweb (serving the production web-site) and Distribution Manager (used for distributing photos to clients).

The photos are stored in their original size in JPEG format, all the text is stored embedded IPTC-IIM and sometimes in XMP.

We are using an automatic cataloguing system to offer categories, categories, geographical places, persons, events and organizations, based on expressions in the description (caption).

The automatic categorization system is based on Oracle database and Orenge knowledge model technology. This technology gives the possibility to translate keywords in the photo metadata, too.

All the metadata is stored in the JPEG file and Fotoware products are able to use it correctly. It gives the possibility to extend the XMP, to create own namespace and define new fields. Of course in that case we have done special services to read these information to serve other applications or to create a sidecar XML format with metadata.

2.7.2 Fields to be filled

During the cataloguing, we try to fill in as many fields as we can.
**XMP schema**

As we mentioned previously we are using the Fotoware software family to manage our photos. From version 6 Fotoware supports XMP. It gives the possibility to manage multilingual fields, and also gives the possibility to define new metadata-fields in an own namespace. This is controlled by metadata schemes, and XMP allows new metadata schemes to be created.

<table>
<thead>
<tr>
<th>field name</th>
<th>value type</th>
<th>max size</th>
<th>IMIT-id (2.xxx)</th>
<th>XMP namespace</th>
<th>XMP tag name</th>
<th>fieldType</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Text</td>
<td>128</td>
<td>5 dc</td>
<td>title</td>
<td>AltLang</td>
<td>Hungarian and English</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Text</td>
<td>3</td>
<td>15 photoshop</td>
<td>Category</td>
<td>Single</td>
<td>(?????)</td>
<td></td>
</tr>
<tr>
<td>Categories</td>
<td>Text</td>
<td>132</td>
<td>20 photoshop</td>
<td>SupplementalCategories</td>
<td>Bag</td>
<td>Hungarian, see &quot;Categories in English&quot;</td>
<td></td>
</tr>
<tr>
<td>Keywords</td>
<td>Text</td>
<td>64</td>
<td>25 dc</td>
<td>subject</td>
<td>Bag</td>
<td>Hungarian, see &quot;Keyword in English&quot;</td>
<td></td>
</tr>
<tr>
<td>Release Date</td>
<td>Text</td>
<td>15</td>
<td>30 fwl</td>
<td>ReleaseDate</td>
<td>Single</td>
<td>Date and time in XMP</td>
<td></td>
</tr>
<tr>
<td>Release Time</td>
<td>Text</td>
<td>15</td>
<td>35 fwl</td>
<td>ReleaseTime</td>
<td>Single</td>
<td>Date and time in XMP</td>
<td></td>
</tr>
<tr>
<td>Special Instructions</td>
<td>Text</td>
<td>256</td>
<td>40 photoshop</td>
<td>Instructions</td>
<td>Single</td>
<td>Hungarian</td>
<td></td>
</tr>
<tr>
<td>Created Date</td>
<td>Date</td>
<td>16</td>
<td>55 fwl</td>
<td>DateCreated</td>
<td>Single</td>
<td>Date and time in XMP</td>
<td></td>
</tr>
<tr>
<td>Created Time</td>
<td>Text</td>
<td>15</td>
<td>60 fwl</td>
<td>CreatedTime</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Text</td>
<td>64</td>
<td>80 dc</td>
<td>creator</td>
<td>Seq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Byline Title</td>
<td>Text</td>
<td>64</td>
<td>85 photoshop</td>
<td>AuthorsPosition</td>
<td>Single</td>
<td>(?????)</td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>Test</td>
<td>32</td>
<td>90 photoshop</td>
<td>City</td>
<td>Single</td>
<td>Hungarian, see &quot;City in English&quot;</td>
<td></td>
</tr>
<tr>
<td>Province State</td>
<td>Text</td>
<td>32</td>
<td>95 photoshop</td>
<td>State</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country Code</td>
<td>Text</td>
<td>3</td>
<td>100 lptphotoCore</td>
<td>CountryCode</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country name</td>
<td>Text</td>
<td>64</td>
<td>101 photoshop</td>
<td>Country</td>
<td>Single</td>
<td>Hungarian, see &quot;Country in English&quot;</td>
<td></td>
</tr>
<tr>
<td>Original Transmission Refer</td>
<td>Text</td>
<td>32</td>
<td>103 photoshop</td>
<td>TransmissionReference</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peoples in the picture</td>
<td>Text</td>
<td>256</td>
<td>105 photoshop</td>
<td>Headline</td>
<td>Single</td>
<td>comma delimited list</td>
<td></td>
</tr>
<tr>
<td>Credit</td>
<td>Text</td>
<td>32</td>
<td>110 photoshop</td>
<td>Credit</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source</td>
<td>Text</td>
<td>32</td>
<td>115 photoshop</td>
<td>Source</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copyright</td>
<td>Text</td>
<td>256</td>
<td>116 dc</td>
<td>rights</td>
<td>AltLang</td>
<td>Hungarian and English (???)</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Text</td>
<td>10000</td>
<td>120 dc</td>
<td>description</td>
<td>AltLang</td>
<td>Hungarian and English</td>
<td></td>
</tr>
<tr>
<td>GPS Latitude</td>
<td>Test</td>
<td>256</td>
<td>exif</td>
<td>GPSLatitude</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPS Longitude</td>
<td>Test</td>
<td>256</td>
<td>exif</td>
<td>GPSLongitude</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original Description</td>
<td>Text</td>
<td>2000</td>
<td>mti</td>
<td>EredetKepszoveg</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hattér információ</td>
<td>Text</td>
<td>2000</td>
<td>mti</td>
<td>HatterInformacio</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPS Longitude (WGS84)</td>
<td>Test</td>
<td>256</td>
<td>mti</td>
<td>GPSStlosessegfok</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPS Longitude (WGS84)</td>
<td>Test</td>
<td>256</td>
<td>mti</td>
<td>GPSStlosessegfok</td>
<td>Single</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Categories in English</td>
<td>Text</td>
<td>132</td>
<td>fwlphotoshop</td>
<td>SupplementalCategories-en</td>
<td>Bag</td>
<td>English, see &quot;Categories&quot;</td>
<td></td>
</tr>
<tr>
<td>Keywords in English</td>
<td>Text</td>
<td>64</td>
<td>fwl</td>
<td>subject-en</td>
<td>Bag</td>
<td>English, see &quot;Keyword&quot;</td>
<td></td>
</tr>
<tr>
<td>Creator in English</td>
<td>Text</td>
<td>64</td>
<td>fwl</td>
<td>creator-en</td>
<td>Seq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>City in English</td>
<td>Text</td>
<td>32</td>
<td>fwlphotoshop</td>
<td>City-en</td>
<td>Single</td>
<td>English, see &quot;City&quot;</td>
<td></td>
</tr>
<tr>
<td>Country name in English</td>
<td>Text</td>
<td>64</td>
<td>fwlphotoshop</td>
<td>Country-en</td>
<td>Single</td>
<td>English, see &quot;Country&quot;</td>
<td></td>
</tr>
<tr>
<td>Peoples in the picture in English</td>
<td>Text</td>
<td>256</td>
<td>fwlphotoshop</td>
<td>Headline-en</td>
<td>Single</td>
<td>English, see &quot;Peoples in the picture&quot;</td>
<td></td>
</tr>
</tbody>
</table>

**2.7.3 The News Agency module**

The News agency module has been developed by agency according to integration software specification in D3.5 document.
After the human translating the pictures are selected, by the given selection criteria. The selected files are moved to a folder to be reachable by the automatic translation module.

In the first step some metadata (keywords, city, AuthorsPosition, rights and Special Instruction fields) are automatically translated. The translation is based on our knowledge base used as a dictionary.

After the successful translation for each picture a couple of file is generated with the same file name and "_.jpg" and "_.xml" extension for bitmap image and metadata correspondingly. During this phase the picture is resized. The JPEG size is defined as 600 pixel on the longer side.

The file name is set to its original MTI specific filename with "MTI_" prefix. These file-pairs are forwarded to the portal via FTP.

![Figure 11: MTI Workflow](image-url)
### 2.7.4 Sample XML file

```xml
<?xml version="1.0" encoding="UTF-8"?>
<EURO-Photo_record>
<header>
<identifier>MTI6BF5663A21B240289E7077B3F275C0D4</identifier>
<datestamp>2011-07-26</datestamp>
<setSpec>EURO-Photo</setSpec>
<setSpec>europeana</setSpec>
</header>
<record
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:photoshop="http://ns.adobe.com/photoshop/1.0/"
xmlns:Iptc4xmpCore="http://iptc.org/std/Iptc4xmpCore/1.0/xmils/"
xmlns:Iptc4xmpExt="http://iptc.org/std/Iptc4xmpExt/2008-02-29/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance/
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<language_independent_metadata>
<Iptc4xmpExt:DigImageGUID>MTI6BF5663A21B240289E7077B3F275C0D4</Iptc4xmpExt:DigImageGUID>
<Iptc4xmpCore:SubjectCode>
  <rdf:bag>
    <rdf:li>15501013</rdf:li>
    <rdf:li>15054000</rdf:li>
    <rdf:li>14002000</rdf:li>
    <rdf:li>03014001</rdf:li>
    <rdf:li>01026501</rdf:li>
    <rdf:li>08005000</rdf:li>
  </rdf:bag>
</Iptc4xmpCore:SubjectCode>
<photoshop:DateCreated>1994-09-11</photoshop:DateCreated>
<dc:creator>Németh György</dc:creator>
<Iptc4xmpCore:CountryCode>HUN</Iptc4xmpCore:CountryCode>
<photoshop:Source>MTI</photoshop:Source>
</language_independent_metadata>
<native_language_metadata>
<dc:language>HUN</dc:language>
<photoshop:headline>Jótékonyság - Angol-magyar labdarúgó mérkőzés</photoshop:headline>
```

Grosics Gyula a korabeli Béke és Szabadság című lapot mutatja, amely a híres 1953-ban játszott 6:3-as magyar győzelemmel zárult angol-magyar mérkőzésről tudósít. Az Aranycsapat egykori tagja a jótékonyság célú angol-magyar labdarúgó mérkőzésen vesz részt, amelyet a várpalotai robbanás sérültjeinek megsegítésére rendeztek.

MTI Fotó: Németh György

Hódmezővásárhely, 10 September 1994


MTI/György Németh
<dc:subject>
  <rdf:bag>
    <rdf:li>articles for use</rdf:li>
    <rdf:li>public figure’s occupation</rdf:li>
    <rdf:li>people</rdf:li>
    <rdf:li>object</rdf:li>
    <rdf:li>newspaper</rdf:li>
  </rdf:bag>
</dc:subject>

<photoshop:City>Hódmezővásárhely</photoshop:City>
<photoshop:State/>

<Iptc4xmpExt:LocationShown/>
<Iptc4xmpExt:PersonInImage>
  <rdf:bag>
    <rdf:li>Grosics Gyula</rdf:li>
  </rdf:bag>
</Iptc4xmpExt:PersonInImage>
<Iptc4xmpExt:OrganisationInImageName>
</Iptc4xmpExt:OrganisationInImageName>

<photoshop:Instructions/>
<photoshop:AuthorsPosition>internal staff</photoshop:AuthorsPosition>
<dc:rights>MTVA - Media Service Support and Asset Management Fund</dc:rights>

</english_language_metadata>
</record>
</EURO-Photo_record>
2.8 PAP local module description & assessment

2.8.1 Technical background
Polish Press Agency is using Fotoware products for managing, archiving and serving photos. The software FotoStation is used by photo editors and people who work for the Photo Archives (including translator) for adding metadata to photos.

Further elements of the Fotoware software family are: Index Manager (used for indexing and searching), FotoWeb (serving the production web-site).

2.8.2 Fields to be filled
During cataloguing we try to fill in as many fields as possible. For the EURO-PHOTO project we provide (O – Optional, M – Mandatory):

<table>
<thead>
<tr>
<th>Label</th>
<th>IPTC Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language independent</td>
<td></td>
</tr>
<tr>
<td>DateCreated</td>
<td>2:055 O</td>
</tr>
<tr>
<td>Byline/Author</td>
<td>2:080 O</td>
</tr>
<tr>
<td>Country Code</td>
<td>2:100 M</td>
</tr>
<tr>
<td>Source</td>
<td>2:115 M</td>
</tr>
<tr>
<td>Copyright Notice</td>
<td>2:116 M</td>
</tr>
<tr>
<td>Polish Language</td>
<td></td>
</tr>
<tr>
<td>Headline</td>
<td>2:105 M</td>
</tr>
<tr>
<td>Keywords</td>
<td>2:025 M</td>
</tr>
<tr>
<td>City</td>
<td>2:090 O</td>
</tr>
<tr>
<td>State/Province</td>
<td>2:095 O</td>
</tr>
<tr>
<td>English Language</td>
<td></td>
</tr>
<tr>
<td>Headline</td>
<td>2:106 M</td>
</tr>
<tr>
<td>City</td>
<td>2:091 M</td>
</tr>
<tr>
<td>Bilingual (PL &amp; EN)</td>
<td></td>
</tr>
</tbody>
</table>
XMP & XML: Polish Press Agency does NOT use xml containing all the metadata for photos, and xmp as well. All the metadata is stored as embedded IPTC-IIM.

2.8.3 The News Agency module

The News agency module has been developed by PAP according to integration software specification in D3.5 document.

The historians are choosing pictures using visualizers. Marked frames are scanned.

**Figure 12: PAP – Historians at work**

Then historians make second selection and provide basic metadata captions. Group of editors make final selection and prepare final metadata in Polish language. The next step is to translate the captions and headlines by group of translators (humans).
Then the photos are moved to the database, and while they enter the database they receive unique filenames.

An editor copies files (already uniquely named) from database to the local folder, changes charset from CP-1250 (that we use in PAP) to UTF-8 and resizes JPEGs to 600 pixel on the longer side, all in FotoStation using macros developed by the IT. Such prepared files are moved via FTP.

**Figure 13 : PAP Screen shoots**
3 Europeana harvesting & online publication.

EURO-Photo acts toward Europeana as an aggregator collecting metadata from a group of content providers (i.e. news agencies) and transmitting them to Europeana. It gathers material from individual organisations, standardise the file formats and metadata, and channel the latter into Europeana according to the Europeana guidelines and procedures. It also supports the content providers with administration, operations and training. The aggregation model is a key element of Europeana Partner Strategy to the development and success of Europeana.

Euro-Photo service is intended to sell pictures to professional users (media market) enabling federated searches in all partners’ archives in a multilingual context that moreover will act as a Europeana aggregator managing an OAI-PMH repository comprising IPTC original metadata data mapped to the EDM for Europeana. Each news agency is from the conceptual point of view a content provider for Europeana providing digital content that are accessible via Europeana. Europeana only ingests and indexes the institution’s metadata while the digital object remains in the news agency repository. Due to the enormous amount of work that the harmonisation and normalisation of metadata requires, Euro-Photo will provide an intermediate layer of content providers acting as aggregators.

![Figure 14: Aggregators in the Europeana organisational model](image-url)
EURO-Photo acts as an aggregator exposing to Europeana an OAI-PMH repository composed by a data set for each Agency. EURO-Photo developed a customized version of an open source implementation of OAI-PMH as described at:
http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

Open source tools list at: http://www.openarchives.org/pmh/tools/tools.php
We selected: http://physnet.unioldenburg.de/oai/
This server support only dublin core standard we adapted it to support Europeana ESE (Europeana Semantic Elements) standard as described at: http://pro.europeana.eu/documents/900548/dc80802e-6efb-4127-a98e-c27c95396d57
OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) server for EUROPEANA feeding.example:


<table>
<thead>
<tr>
<th>Data collection</th>
<th>Published at</th>
</tr>
</thead>
<tbody>
<tr>
<td>All archives together</td>
<td>EURO-Photo portal (<a href="http://www.europhto.eu.com">www.europhto.eu.com</a>)</td>
</tr>
<tr>
<td>Agencies’ historical archives</td>
<td>Agency’s web site</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data collection</th>
<th>SetSpecs</th>
</tr>
</thead>
<tbody>
<tr>
<td>EuroPhoto – ANSA</td>
<td>&lt;setName&gt;ANSA&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto – ANP</td>
<td>&lt;setName&gt;ANP&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto - BELGA</td>
<td>&lt;setName&gt;BELGA&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto - DPA</td>
<td>&lt;setName&gt;DPA&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto - EPA</td>
<td>&lt;setName&gt;EPA&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto - EFE</td>
<td>&lt;setName&gt;EFE&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto - PAP</td>
<td>&lt;setName&gt;PAP&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto – LUSA</td>
<td>&lt;setName&gt;LUSA&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto – MTI</td>
<td>&lt;setName&gt;MTI&lt;/setName&gt;</td>
</tr>
<tr>
<td>EuroPhoto - SCANPIX</td>
<td>&lt;setName&gt;SCANPIX&lt;/setName&gt;</td>
</tr>
</tbody>
</table>
Figure 15: EURO-Photo pictures shown at Europeana
Figure 16: EURO-Photo picture shown at Europeana
Figure 17: Europeana links EURO-photo (ShownAt mode)
Photos transferred to Europeana at the end of the project:

<table>
<thead>
<tr>
<th>Agency</th>
<th>Uploaded to Europeana</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ANSA</td>
<td>75,507</td>
</tr>
<tr>
<td>2 ANP</td>
<td>155,448</td>
</tr>
<tr>
<td>3 Belga</td>
<td>37,616</td>
</tr>
<tr>
<td>4 DPA</td>
<td>132,175</td>
</tr>
<tr>
<td>5 EFE</td>
<td>118,858</td>
</tr>
<tr>
<td>6 EPA</td>
<td>20,356</td>
</tr>
<tr>
<td>7 LUSA</td>
<td>31,162</td>
</tr>
<tr>
<td>8 MTI</td>
<td>120,870</td>
</tr>
<tr>
<td>9 PAP</td>
<td>119,548</td>
</tr>
<tr>
<td>10 Scanpix</td>
<td>97,269</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>908,809</strong></td>
</tr>
</tbody>
</table>
CONCLUSION

This document summarizes the final assessment activities of EURO-Photo portal prototype (www.europhoto.eu.com), describes News Agencies local modules implementation & assessment and Europeana harvester implementation and upload activities.

The result of the assessment activities is positive.
APPENDIX I: Europeana content provision Form