



ICT-PSP Project no. 297158

EUROPEANAPHOTOGRAPHY

EUROPEAN Ancient PHOTographic vintaGe repositoRies of digitAized Pictures of  
Historical qualitY

Starting date: 1stFebruary 2012

Ending date: 31stJanuary 2015

<b>Deliverable Number:</b>	D 4.1
<b>Title of the Deliverable:</b>	EuropeanaPhotography Vocabulary Definition
<b>Dissemination Level:</b>	Restricted

<b>Contractual Date of Delivery to EC:</b>	Month 12
<b>Actual Date of Delivery to EC:</b>	31 January 2013

Project Coordinator

*Company name :* KU Leuven  
*Name of representative :* Fred Truyen  
*Address :* Blijde-Inkomststraat 21 B-3000 Leuven PB 3301  
*Phone number :* +32 16 325005  
*E-mail :* fred.truyen@kuleuven.be  
*Project WEB site address :* <http://www.europeana-photography.eu>

Technical Coordinator

*Company name :* Promoter S.r.l.  
*Name of representative :* Antonella Fresa  
*Address :* Via Boccioni, 2 I-56037 Peccioli (Pisa)  
*Phone number :* +32 0587 677724  
*E-mail :* fresa@promoter.it  
*Project WEB site address :* <http://www.europeana-photography.eu>

## Context

WP 4	Indexing and multilingual support
WPLLeader	KMKG – Koninklijke Musea voor Kunst en Geschiedenis
Task 4.3-4.4	EuropeanaPhotography Vocabulary Definition & Vocabulary Translation
Task Leader	KMKG
Dependencies	WP3, WP5, WP7

Author(s)	KMKG
Contributor(s)	All the partners
Reviewers	Frank Golomb (United Archives), Hanne Abildgaard (Arbejdermuseet)
Approved by:	Antonella Fresa, Technical Coordinator

## History

Version	Date	Author	Comments
0.1	24/01/2013	KMKG	
0.2	29/01/2013	KMKG	Modifications after the internal review
0.3	30/01/2013	Promoter	Global reading and graphic adjustments
1.0	31/01/2013	Promoter	Final version

## TABLE OF CONTENTS

1	Executive Summary	4
2	Introduction	5
2.1	Background	5
2.2	Role of this deliverable in the project	5
2.3	Approach	5
3	Sources	7
3.1	AAT	7
3.2	IPTC Information Interchange Model	7
3.3	MARC 21	7
3.4	SEPIADES	8
3.5	Partners' own vocabularies	8
4	Structure of the vocabulary	9
4.1	Photographic Techniques	9
4.2	Photographic Type/Practice	9
4.3	Keywords	9
5	Further developments	10
6	Conclusions	11

## 1 Executive Summary

While all content for the EuropeanaPhotography project is similar in technique and date, there is an enormous variety in subject matter, ranging from nature photography to press photography all over Europe. The variety is substantial in most of the content partners' collections, so it stands to reason that the consortium as a whole finds a way to combine metadata from all partners in order to simplify searches, enrich search results and inscribe the consortium's metadata within the whole of Europeana.

We've chosen to do this by creating a multi-faceted, hierarchically structured, 12-language vocabulary, covering a large number of subjects, photographic types and photographic techniques. It is based on internationally recognised standards and utilises the knowledge and expertise present in the content partners' institutions.

This deliverable describes the structure of the vocabulary, its sources, and the workflow used in order to achieve the necessary results, as well as the use of the vocabulary within the project, the plans for dissemination of the vocabulary and the continuation of the work during and after the end of the EuropeanaPhotography project.

The document is composed by 6 chapters:

Chapter 1 Executive Summary

Chapter 2 Introduction – this chapter describes the scope of the work done for the development of this deliverable and the approach taken

Chapter 3 Sources – this chapter describes the sources used for the compilation of the vocabulary, namely AAT, IPTC, MARC21, SEPIADES, Partners' own vocabularies

Chapter 4 Structure – this chapter describes the structure of the EuropeanaPhotography multilingual vocabulary which includes photographic techniques, photographic types and practices, keywords belonging to the photographic sector

Chapter 5 Further developments – this chapter describes the future activities which could be beneficial to improve the vocabulary

An excel file is attached to this document. It contains the words in the vocabulary and their translations in 12 languages.

## 2 Introduction

This document provides a view of the EuropeanaPhotography vocabulary, from its creation (necessity, sources (chapter 3), structure (chapter 4) to its use within, throughout and after the project (chapter 5).

### 2.1 Background

While all photography is reasonably similar in and restricted by technique, especially when reduced to a limited timeframe, its subject matter is enormously diverse. Furthermore, the context and the reason behind the creation of a certain photography can be more telling to historians, researchers and the general public than the image itself. This in itself is sufficient reason for the creation of a hierarchically structured vocabulary of techniques, keywords and photographic types to help all interested parties find the information they require.

Within the EuropeanaPhotography project, partners have selected photographs from their collections dating from the earliest photographs to the beginning of the second World War. While anyone was free to make the choices they desired, certain recurrent themes became evident in all collections. However, since all partners describe their photos in their own language, there was no way to link related photos to each other.

Both reasons meant that the creation of a simplified but uniform, multilingual<sup>1</sup> vocabulary for this project was an excellent tool in the enrichment of the project's metadata delivery to Europeana. This vocabulary will allow fine-grained navigation thanks to its hierarchical structure, supports the multilingual entries by the content partners and can be easily exported in SKOS.

### 2.2 Role of this deliverable in the project

The vocabulary defined within this deliverable is a key element of the semantic enrichment of the partners' metadata. While not strictly necessary for the digitization of photography or delivery to Europeana, this vocabulary ensures that themes can be traced throughout the different collections. It provides translations for the keywords associated with photographic subjects and techniques, and allows better and more complete search results when searching Europeana for photographic content.

The vocabulary will support WP3 – Digitisation – through simple integration with the MINT2 mapping tool provided by technical partner NTUA, and will be used by WP7 – Dissemination – in the creation of thematic expositions and other dissemination activities.

### 2.3 Approach

The creation of the EuropeanaPhotography vocabulary was started by a survey of language features. In this survey, all partners were requested to state, among others, what vocabularies they were using (see also 3: sources), in which ways these vocabularies were used and to share their vocabularies. This survey was completed in the early months of this project and the

---

<sup>1</sup> In no less than 12 languages: English (as the pivotal language), French, Dutch, German, Italian, Polish, Danish, Bulgarian, Slovak, Lithuanian, Catalan and Spanish



results communicated to the consortium at the content seminar in M3. The results have also been included in the content seminar proceeding, D2.1.

At this same content seminar, 3 main facets of the vocabulary have been decided upon (see 4: structure). From that point onwards, several drafts of the vocabulary have been made, shared with the content partners, discussed and refined.

After a final English version was agreed, translations were started by the partners as foreseen in the DoW, resulting in the multifaceted, hierarchically structured, nearly 500-concept multilingual vocabulary provided for you today.

### 3 Sources

Five kinds of sources have been considered in the design of the EuropeanaPhotograph Vocabulary, i.e.:

1. AAT – Art and Architecture Thesaurus Online, the vocabulary of the Getty Research institute, <http://www.getty.edu/research/tools/vocabularies/aat/>
2. IPTC IIM – Information Interchange Model of the International Press Telecommunications Council, <http://www.iptc.org/site/index.html?channel=CH0100>
3. MARC 21 – The format for Bibliographic Data, <http://www.loc.gov/marc/bibliographic/>
4. SEPIADES – The recommendations for cataloguing photographic collections of the International Council of Archives, <http://www.ica.org/7363/paag-resources/sepiades-recommendations-for-cataloguing-photographic-collections.html>
5. Partners' own vocabularies

#### 3.1 AAT

The AAT is a faceted controlled vocabulary used for describing items of art, architecture and material culture. AAT contains generic terms and is used in a wide range by museums, libraries and archives. AAT is also used in EuropeanaLabs, as it is one of the rare examples of relevant polyhierarchical thesauri in the field of cultural heritage. The AAT is built in compliance with the ISO 2788 standard and it contains over 130.000 terms, and it is maintained by the Getty Research Institute. For EuropeanaPhotography, we are mainly interested in the photographic techniques describes herein. They are however described too extensively for use in this project, and used as inspiration only.

#### 3.2 IPTC Information Interchange Model

Many of the project's content partners use this model that has been specifically designed for the exchange of news among international papers and agencies. The IPTC model has been used worldwide by professional photographers and photographic agencies for embedding metadata in a structured way in the photographs.

Most of our partners use extended or adapted IPTC headers in their databases, depending on their needs.

#### 3.3 MARC 21

Developed by the Library of Congress in the 1960s for bibliographic items, the MARC standard, and in particular MARC21, have also been used by many of our content partners. MARC



standards use authority records for several fields of the metadata description. When used by our partners, often the lists have been adapted for internal use.

MARC is based on [ISO 2022](#) and allows the use of Hebrew, Cyrillic, Arabic, Greek, and East Asian scripts. MARC 21 in UTF-8 format allows all the languages supported by Unicode.<sup>2</sup>

### **3.4 SEPIADES**

Sepiades is a data set specifically designed to describe photographic objects, developed in the EU-funded SEPIA project. It allows multilevel description on techniques, keywords and other subsets, makes a distinction between objects themselves and collections, and has a software tool for implementation of the model.

### **3.5 Partners' own vocabularies**

Nearly all partners have developed, more or less extended, their own vocabularies, often based on one of the before mentioned standards. These have been used as the primary source for the creation of the EuropeanaPhotography vocabulary, as they are best adapted to the content partners' needs.

---

<sup>2</sup> <http://www.loc.gov/marc/specifications/speccharmac8.html>



## **4 Structure of the vocabulary**

The EuropeanaPhotography consortium has chosen to use a modular structure for the thesaurus to keep a good overview regarding the concepts and their arrangement in the hierarchical, relational structure. This structure has been agreed at the content seminar in M3, and refined throughout the course of the development of the vocabulary.

### **4.1 Photographic Techniques**

The first part of the thesaurus lists, in a hierarchical way, are photographic techniques used to create the object. It mainly distinguishes between positive and negative types of photography. The main sources for this part of the thesaurus are AAT and SEPIADES. The techniques are linked to the development process of photography at the time, and therefore limited to the techniques available between the first photographs and the second World War.

### **4.2 Photographic Type/Practice**

In this part the consortium has tried to capture the reason for taking a specific photograph. The photographic practice describes in keywords, of which one or more can be selected, the way a photo was conceived and for which goal was intended. This goal often holds information that is not immediately visible on the actual picture itself, but is necessary or interesting for research.

### **4.3 Keywords**

In the keywords section a large number of subjects have been regrouped for the content partners to describe the subject of their photographs. This section is mainly based on the IPTC model, SEPIADES and the partners' vocabularies. While extensive, this list of keywords is a consensus between the participants and a simplified version of their vocabularies. It will be used most extensively in the enrichment of metadata and translation for publication on Europeana.

## 5 Further developments

The current version of the thesaurus, nearly 500 concepts, already allows the content partners to annotate, translate and semantically enrich their metadata. More work is expected to be done in the frame of the cooperation with other projects, to expand the possibilities for semantic enrichment, mainly by creating possibilities to link metadata to, for instance, GeoNames. To this extent all content partners will deliver input based on their own needs and from their own fields of expertise. The semantic enrichment will progress by further developing and incorporating structured lists of events, individuals and locations.

Also, in the framework of the technical, financial and timing possibilities, efforts will be made to constantly improve the semantic richness of the thesaurus. Methods to be used to this extend will be considered on an ad hoc base and communicated with all partners before implementation.

In the light of possible future work, it is worth to mention that EuropeanaPhotography project is currently establishing cooperation agreements with other projects, with particular regard to Linked Heritage and EuroPhoto.

Linked Heritage ([www.linkedheritage.eu](http://www.linkedheritage.eu)) is a Best Practice Network with a broad spectrum of goals; one of them is the objective to demonstrate enhancement of quality for the content provided to Europeana, in terms of metadata richness, re-use potential and uniqueness. As KMKKG, the WP Leader responsible of this Deliverable, is involved also in Linked Heritage as WP Leader about multilingualism, the cooperation between the two projects is absolutely natural and can be very productive.

An interesting dialogue has been established with EuroPhoto (<http://www.europhoto.eu.com>), a digitisation Pilot B project that is completing the digitisation of 150,000 images from 10 leading European News Agencies. The cooperation between EuroPhoto and EuropeanaPhotography will continue also in the domain of the multilingual terminologies for photographic digital content.

The EuropeanaPhotography vocabulary will be presented on the project's website ([www.europeana-photography.eu](http://www.europeana-photography.eu)) and on the showcase of EuropeanaPhotography realised on digitalmeetsculture.net portal (<http://www.digitalmeetsculture.net/heritage-showcases/page-2/>).

## 6 Conclusions

This nearly 500-word vocabulary has been an important undertaking to help unify all content in the project coming from the different partners. It allows not only translation and semantic enrichment of the metadata at the time of publication on Europeana, but helps guide all partners in the metadata creation.

Being the EuropeanaPhotography consortium very much heterogeneous, as it is composed of very different partners with very different experience, requirements and needs, it was necessary to find a consensus and a sort of compromise among them all. For example, in the consortium there are partners with limited to no structured lists of terminology, and partners with thousands of terms which go into enormous detail. For this reason, there was an evident gap between expectations, wishes and use among the partners. It wasn't feasible in the scope of this project to recreate a vocabulary with the same wide variety of keywords as some of the partners have, in particular the commercial partners, or the extensive level of depth in a limited thematic field, such as other partners. This project, and this work package, has decided to make a vocabulary that brought the greatest use to the greatest amount of partners.

After the initial analysis, it came to evidence that for the 'average' EuropeanaPhotography partner there is no need to have a too extensive list of keywords, while on the other hand the vocabulary needed to include a lot of topics. Furthermore, in some cases not all topics related to a photo are known yet. Therefore, it was not possible to simply implement an existing set of keywords, but it was necessary to develop a specific one. This task was quite demanding indeed, and had to be done in a limited timeframe between the completion of partners' language survey and the allowed time necessary for translations. Anyhow the result is very satisfying and the achievement is a good balance between having a wide range of topics and a good depth of vocabulary.

Throughout the next phase of the project, the vocabulary will be used for semantic enrichment in the description of the photographs delivered to Europeana, by adding the keywords in the mapping stage through the MINT-tool. It will allow users to consult all collections at once with the use of keywords in their own language, and will unify the content from the different partners among thematic lines. Furthermore, it will serve as a way of creating thematic views of the content to be delivered on the Europeana Photography website for dissemination.

## **ANNEX I: THE EUROPEANAPHOTOGRAPHY VOCABULARY**

The whole vocabulary is annexed to this document as an excel file.