DELIVERABLE

Project Acronym:	**LINKED HERITAGE**
Grant Agreement number:	**270905**
Project Title:	**Coordination of standard and technologies for the enrichment of Europeana**

D 7.3 Training Materials

Revision: version 2.0

Authors:

Antonella Zane (UNIPD)
Elena Bianchi (UNIPD)
Gianluca Drago (UNIPD)

Contributors:

Luisa Banzato (UNIPD)
Cristiana Bettella (UNIPD)
Lisa Dainese (UNIPD)
Gianluca Drago (UNIPD)
Michele Visentin (UNIPD)

Reviewers:

Maria Teresa Natale (ICCU)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
0.1	22/3/2013	Antonella Zane Elena Bianchi Gianluca Drago	CAB/UNIPD	First draft
0.2	30/03/2013	Antonella Zane Elena Bianchi Gianluca Drago	CAB/UNIPD	Second draft
0.3	01/04/2013	Maria Teresa Natale	ICCU	Peer review
0.4	04/04/2013	Antonella Zane Elena Bianchi Gianluca Drago	CAB/UNIPD	Third draft after peer review
0.5	05/04/2013	Antonella Zane Elena Bianchi Gianluca Drago	CAB/UNIPD	Final version
1.0	07/04/2013	Claudio Prandoni	PROMOTER	Formal check
1.1	14/09/2013	Antonella Zane Elena Bianchi Gianluca Drago	CAB/UNIPD	First draft of final deliverable
1.2	26/09/2013	Maria Teresa Natale	ICCU	Peer review
1.3	27/09/2013	Antonella Zane Elena Bianchi	CAB/UNIPD	Final version
2.0	30/09/2013	Claudio Prandoni	PROMOTER	Formal check

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

EXECUTIVE SUMMARY

This document is the **final version of D7.3 Deliverable**. It integrates and substitutes the D7.3 intermediate version released on month 24 as scheduled by the Linked Heritage Document of Work (DOW).

In addition, we would like to emphasise that **D7.3 is a compendious description of the Training materials realised within the project or collected from authoritative sources and systematised for learning purposes by the University of Padova (UNIPD) in collaboration with the Linked Heritage e-learning Working Group. The training materials (Learning objects, booklets, guidelines, case studies, bibliographies, etc,) are available in the web at:**

https://elearning.unipd.it/cab/course/view.php?id=4; http://linkedheritage.cab.unipd.it/training/LO-00/en/overview.html.

Furthermore, Learning objects and the Training programme overview are also available in printable PDF files in Annex II.

D7.3 Training materials (focused on the content of the LH comprehensive training programme) is complementary to D7.4 Virtual Learning Environment (focused on the web platform which hosts the comprehensive training programme). This is why we invite to read both the document in order to have the overall picture of the activities carried out and the results obtained by the Task Training of the WP7.

WP7 Task 7.2-Training complements Task 7.1-Dissemination and works on the results and outputs of all the other Work Packages:

- WP2: Linking Cultural Heritage information (last deliverable due by M24)
- WP3: Terminology (last deliverable due by M24)
- WP4: Public Private Partnership (last deliverable due by M24)
- WP5: Technical integration
- WP6: Coordination of content

As the task training of the **WP7 elaborates and is strictly connected to the outputs of all the other Linked Heritage Work Packages**, both the final versions of these Deliverables are released by the University of Padova, on behalf of the Linked Heritage project, on Month 30, and not on month 24 as planned by DOW.

The D7.3 deliverable provides a **detailed description of the Linked Heritage Comprehensive Training Programme** (Syllabus in DOW) and of the open educational content that the University of Padova has accomplished.

This deliverable is divided into two parts:

a) a description of the work done (chapters 1 - 6)

b) the image of the actual deliverable: PDF files representing all the Learning Objects webpages, the associated readings and training materials and the questionnaires (evaluation and the assessment) captured by Print pages to Pdf (Annex 2).

The first part of the deliverable is structured as follows:

- Chapter 1 contains the background and an overall description of the Linked Heritage project in terms of main goals, expectations and approach. It introduces the tasks WP7.1 Dissemination and WP7.2 Training.

- Chapter 2 outlines the Comprehensive Training Programme (Syllabus) in detail
- Chapter 3 deals with all the issues associated with authoritative educational content - selected and collected with the support of LH partners - and with the process of creating new learning resources.
- Chapter 4 lists the activities carried out by University of Padova to promote Linked Heritage Training Materials.
- Chapter 5 summarises the next steps and conclusions.

The deliverable also includes a bibliography and two Annexes:

- Annex 1 containing a list of videos used in the Learning Objects
- Annex 2 providing screenshots of the Learning Objects .

1 INTRODUCTION

1.1 BACKGROUND

Linked Heritage is a project funded by the European Union as part of its Competitiveness and Innovation Programme Framework Programme (CIP, 2007 – 2013). The project began on 1 April 2011 and ends on 30 September 2013. It involves 38 partners from 20 countries, including EU member states, Israel and Russia.

1.2 THE MAIN GOALS OF THE PROJECT

Linked Heritage has 3 main goals:

- to contribute large quantities of new content to Europeana, from both public and private sectors;
- to enhance the quality and richness of Europeana metadata;
- to improve the functionalities available for search, retrieval and use of Europeana content.

1.3 EXPECTATIONS

The project seeks to expand and enrich Europeana content and services by:

- providing access to 3 million new content;
- facilitating the comprehension and diffusion of key Digital Library concepts by creating new dissemination, information and training tools (websites, posters, booklets, leaflets and learning objects);
- developing new, free software web applications to facilitate the flow of data to Europeana.

1.4 ROLE OF THE UNIVERSITY OF PADOVA LIBRARY CENTRE IN THE PROJECT

The University of Padova Library Centre (CAB) is the Leader of Work Package 7, which includes Task 7.1 “Dissemination” (whose activities are carried out by the ICCU) and Task 7.2 “Training”. The Centre coordinates the Linked Heritage e-learning Working Group that was created to support training activities - the training and the dissemination of results achieved by the Linked Heritage (LH) project. It also

- draws up a training programme focused on the key aspects of the project (Europeana, metadata standards, linked data, persistent identifiers, multilingual terminologies, public-private partnerships);
- designs and produces a set of learning objects in collaboration with the participants of the Linked Heritage Working Group on e-learning;
- creates a Virtual Learning Environment (VLE) devoted to the Linked Heritage training programme.

1.5 MAIN GOALS OF WP7

They can be summarised as follows:

- to increase the size of the Best Practice Network by attracting new members;
- to stimulate the contribution of content to Europeana by raising awareness of the tools, facilities and best practice provided by Linked Heritage;
- to raise awareness of the Linked Heritage work across the Europeana ecosystem, and to encourage Europeana itself, as well as content providers and aggregators, to take full advantage of the outcomes of the project;

- to build stronger links between public and private sectors;
- to build technical skills and knowledge in the cultural heritage sector, especially in terms of Europeana and Linked Heritage technologies
- to create, deliver and publish reusable Learning Objects (LOs) and other training materials and resources designed and developed for multiple learning approaches.

1.6 WP7 TASK DISSEMINATION: GOALS AND TARGET AUDIENCE

WP7 has two main tasks: dissemination and training. Before analysing the WP7.2 Training, which is the subject of this deliverable, we would like to remind readers of the main goals of the Linked Heritage WP7.1 dissemination activities, which are to

Raise awareness, Inform, Engage and Promote.

The target audience of Linked Heritage is very specific and includes the following categories: the community of content providers and aggregators, government and policy bodies, the private (publishing) sector, digitisation professionals, libraries, archives, museums, the research community and Europeana Labs (see D.7.5 § 3.8).

1.7 WP7 TASK TRAINING: GOALS AND ACTIVITIES

As stated by the Description of Work (DoW), the main goals of WP7.2-Training are to design, develop and implement a comprehensive training syllabus of the most important project topics. In this deliverable, we also refer to the "comprehensive training syllabus" as a "comprehensive training programme".

WP7.2 main activities can be summarised as follows:

- to gather training materials from different authoritative sources; these will also include the results of the experimental work done by the WP2 and WP3;
- to plan and produce a set of new Learning Objects;
- to disseminate all the materials both through Web channels and through a Virtual Learning Environment.

1.8 OUR APPROACH

The development of a training programme requires in-depth knowledge of the targeted users' needs and a clear identification of the intended learning outcomes. This is why we deemed it crucial to involve in the Task activities Linked Heritage partners from a wide range of user categories. As regards task 7.2-Training, we set up a thematic Working Group (WG) on e-learning that acted as a Lab in which participants contributed to the elaboration of the training programme and to the development of learning objects, tested the first "prototypes", translated content and provided a continuous feedback on the effectiveness of the results available in the Virtual Learning Environment.

As we believe that the availability of contents in different languages enhances dissemination to potentially interested audiences, partners interested in providing translations of learning objects for their native-speaker users were warmly invited to join the Working Group to have a closer look at the progress of our work and to prioritise the learning objects to be translated for their audience.

In addition, we thought it was important for WP 2-6 Leaders and Linked Heritage specialised professionals to contribute to learning objects content curation, to provide selected readings and training materials, and to review the definitions of words in the Linked Heritage Moodle Glossary. The glossary has been designed in order to provide a sort of conceptual map devoted to the main themes developed in the contest of the Linked Heritage Project. Furthermore it stands as a quick device to better support users approaching first the Linked Heritage Virtual Learning Environment as well as in the learning process.

As regards the target users of the Training programme, they were initially identified in institutions contributing to Europeana as data providers or aggregators, as stated in the DoW (Annex I) of Task 7.2 Training - "Linked Heritage will offer a comprehensive training syllabus to content providers and aggregators, both those who are using Linked Heritage technologies and those who are having difficulties with contributing content to Europeana".

At the kick-off meeting held in Rome in April 2011, partners proactively discussed the scope of the programme, the contents to be included and the intended audience for the educational resources being created. We carefully considered our partners' feedback. Indeed, the current comprehensive training programme stems from all the proposals presented and discussed during the plenary and technical meetings and agreed upon with the WP7 e-learning Working Group[1].

1 Kick-off meeting, Rome, 29 April 2011, http://www.linkedheritage.eu/getFile.php?id=109

2nd plenary and WG meetings, Barcelona, 23-24 November 2011, http://www.linkedheritage.eu/getFile.php?id=217

3rd plenary and WG meetings, Stockholm, 24-25 May 2012, http://www.linkedheritage.eu/getFile.php?id=308

4th plenary and WG meetings, Lisbon, 29-30 November 2012, http://www.linkedheritage.eu/getFile.php?id=416

5th plenary and WG meetings, Padova, 6-8 March 2013, http://www.linkedheritage.eu/getFile.php?id=446

6th plenary meeting, Dublin, 18 June 2013, http://www.linkedheritage.org/getFile.php?id=463

2 THE COMPREHENSIVE TRAINING PROGRAMME

The programme was drawn up with the essential support and contribution of the Linked Heritage e-learning Working Group participants.

As for the Linked Heritage training programme, the following four user groups were identified as representative of the specific category of users:

A. cultural institution managers and decision-makers as potential Europeana content providers:

B. teachers, educators, scholars who could benefit from the exploitation of Europeana content;

C. Library and Information Science (LIS) professionals, entry-level students and museum, library and archive technicians who need to be constantly up-to-date with the development of digital libraries (self-taught and lifelong learners);

D. market players from the private sector.

At the first e-learning Working Group meeting held in Barcelona on 24 November 2011, UNIPD presented the first draft of the structure of the comprehensive training programme.

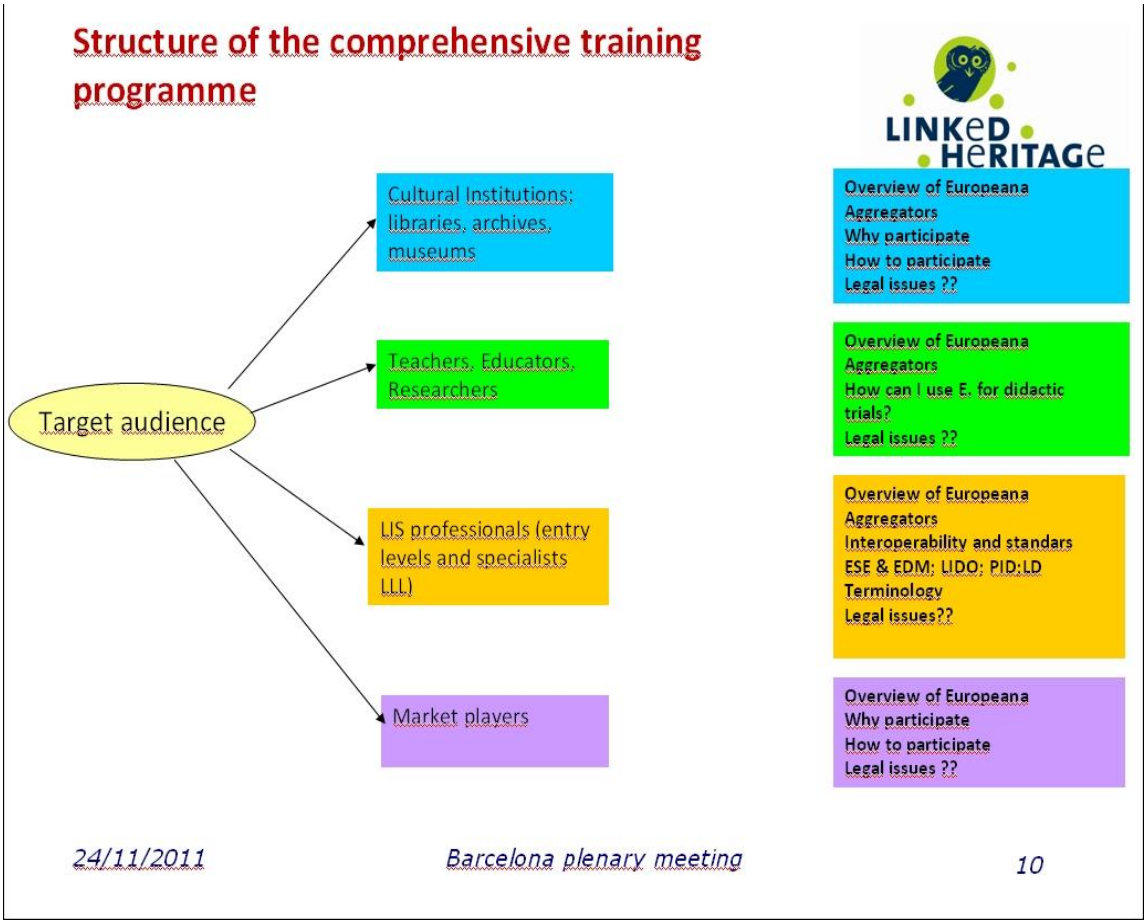


Figure 1 - the first draft of the structure of the Linked Heritage comprehensive training programme

In Barcelona the e-learning Working Group participants were divided into 4 groups according to the structure in figure 1. The groups were either made up of participants from specific categories of users (Cultural Institutions Decision Makers, Teachers and Researchers, Library and Information Science Professionals, Market Players) or were people who could identify the users' needs. Each group envisaged possible users' scenarios, discussed their training needs and prioritised the Learning objects to be developed.

In May 2012, at the second e-learning Working Group meeting in Stockholm, UNIPD described the revised version of the "Structure of the comprehensive programme" (v.2.0) including the participants' suggestions on the topics to be included in the programme:

A. Cultural institutions decision makers

- Aggregation;
- Why participate & How to participate;
- Overview of Europeana (general context, the European Commission);
- Policies (Comité des sages report, Recommendation of the 28th October 2011, etc.);
- Legal issues (Linked Heritage in the Europeana context, presentation of the Linked data).
- Persistent identifiers (included by UNIPD)

B. Teachers and researchers

- General overview of Europeana (specific presentation of the content of Europeana, with regards to research programmes and research fields);
- Tutorial for searching in Europeana;
- User scenarios for educational purposes;
- Information on legal issues: how to re-use contents

C. LIS professionals

- Ontology;
- URI;
- RDF;
- API;
- Data quality – consistency;
- XML;
- Data reuse (in practice): APIs, Mash-ups, Linked data, Data visualisation;
- SKOS;
- Search techniques;
- Digitisation life cycle;
- Case studies on the subjects
- Persistent identifiers (included by UNIPD)

D. Market players

- Overview of Europeana;
- Why participate & How to participate ;
- Legal issues;
- Interoperability and standards.

In the following months the structure of the programme evolved, also thanks to the results of the project, and UNIPD reviewed it to refine and categorise the list of topics.

The **latest version** of the training programme was presented at the Linked Heritage events held in Dublin on 18 June 2013. It is made up of **four modules** specifically targeted to the user groups A, B, C, D. The modules provide independent learning paths, and individuals can choose the topics they are most interested in.

Each module contains a **set of topics** suitable for each target group of users:

A. Cultural institutions decision makers

- Overview of Europeana
- Persistent identifiers
- Digitisation life cycle
- Why and how to contribute to Europeana

B. Teachers and researchers

- General overview of Europeana
- Tutorial for searching on Europeana
- Users scenarios for educational purposes
- Legal issues: how to reuse contents?

C. LIS professionals

- Persistent identifiers
- Persistent identifiers: commercial and heritage views
- Digitisation life cycle
- Terminology
- Linked data
- MINT services

D. Market players

- Overview of Europeana
- Persistent identifiers: commercial and heritage views
- Public-private partnerships with Europeana
- Why & How to participate
- Persistent identifiers

Each topic features a learning package that includes: Case studies, Training materials and Bibliography. Some topics also include a multimedia Learning object that comes with two questionnaires (assessment and evaluation) and further important information: summary and target audience (About the Learning object), technical requirements (Using the Learning object), and aims (To provide skills).

3 TRAINING MATERIALS

What does it training materials mean? “Materials used by instructors, facilitators, and students in a training environment. These materials store the information needed by the learner to perform tasks.” http://wiki.answers.com/Q/What_is_the_definition_of_training_materials

In the Linked Heritage Virtual Learning Environment the term “training material” encompasses four different types of resources:

- **Learning objects**
- **Training materials in the strict sense**
- **Case studies**
- **Bibliography**

3.1 LEARNING OBJECTS

They have been defined as “small digital lesson units that can be reused and combined to create customised courses for lessons”[2].

Learning Objects can have one or more learning objectives, and may contain different digital resources (texts, audio, videos, animations, images). Learning objects are usually managed and accessed through a Virtual Learning Environment.

We presented our first pilot Learning object “Persistent Identifiers: What if?” at the Stockholm plenary meeting (May 2012). The pilot Learning object is an animated video with cartoon-style text balloons and sound effects. The protagonists are two owlets taken from the Linked Heritage logo. The Flash animation was developed by CMELA, Centro Multimediale e di E-Learning di Ateneo of the UNIPD.

Figure 2 – The comic strip of the first pilot Learning object “Persistent Identifiers: What if?”

2 European Commission (2006), *Information Society and Education: Linking European Policies*, http://www.fp7.org.tr/tubitak_content_files/308/dokuman/education.pdf

Based on the usability results of the pilot Learning object and on feedback from the e-learning Working Group, a new, HTML-based Learning object layout was designed to provide users with an interactive approach, to be easily translated into the partners' native languages and to facilitate the reuse and access of Linked Heritage Learning objects not only through the Virtual Learning Environments but also through web sites and multimedia web channels.

The new layout, called "Linked Heritage Learning Objects", was adopted as a template for the other Learning objects.

Figure 3 – The Linked Heritage Learning Objects HTML template

3.1.1 How we implemented the Linked Heritage Learning Objects template

Objectives

After analysing the characteristics of the website – purpose, content and target – we set ourselves the following objectives:

- To create a clearly identifiable straightforward standalone environment suitable to accommodate the Learning objects, their texts, images, external links and videos
- That allows Learning objects to be divided into smaller units linked by sequential progression buttons
- That enables the reuse of Learning objects in different contexts
- That allows the implementation of translations into other languages.

Page layout

The page features a fixed layout (the classic 960px wide) divided into two columns. The width of the right column is defined by the need to include videos. The richness of content in some areas is counterbalanced by the insertion of large gaps.

The separation of chapters inside the page is clearly marked by the graphics used as background. From top to bottom, these are the areas:

- **Header**, including the logotype carrying the site name
- **Navigational buttons**
- **Title of current Learning Object**
- **Title of current page**
- **Left column**. It includes auxiliary links to *Reading and training materials* on the Virtual Learning Environment and, in the presence of videos, the texts accompanying them, the transcripts of speeches and a link to high-definition versions
- **Right column**. it contains the main content of the Learning Object
- **Navigational buttons**
- **Footer**, with logos of project contributors

Figure 4 – the Linked Heritage Learning Object Layout

Style, fonts, colours and graphics

The style is light and bright. The choice of colours, graphic components, the spatial arrangement of the objects and the size of the empty space all aim to create a relaxing but sharp view. The left alignment of components, use of sans-serif font, soft shading and web-fonts for the headlines, impalpable textures in the background and delicate reflections in the graphic panes give it an up-to-date feel. To be consistent with the colours used in the Linked Heritage logo and booklets, we chose a limited colour palette. Graphics are rendered well and are not overwhelming. They may have ornamental functions or emphasise the page structure and its individual elements, or they underline the meaning of certain links. In addition to the main CSS, there is also a style optimised for printing.

Directory structure

Since Learning objects have to be easily adapted to other contexts, we created a specific directory for each one of them, including all the files required for proper web display: HTML, CSS, JavaScript, accompanying texts (video transcripts and subtitles), as well as high-definition images and videos. Any translation of the Learning objects is contained in its own subdirectory.

Additional notes

All the Learning objects are and will be hosted and maintained, even after the end of the Linked Heritage project, by the server farm of the UNIPD Library Centre (CAB). All the embedded videos (45 so far) are hosted under their own account on YouTube, which offers streaming services not yet available at CAB (see D7.4, fig. 14).

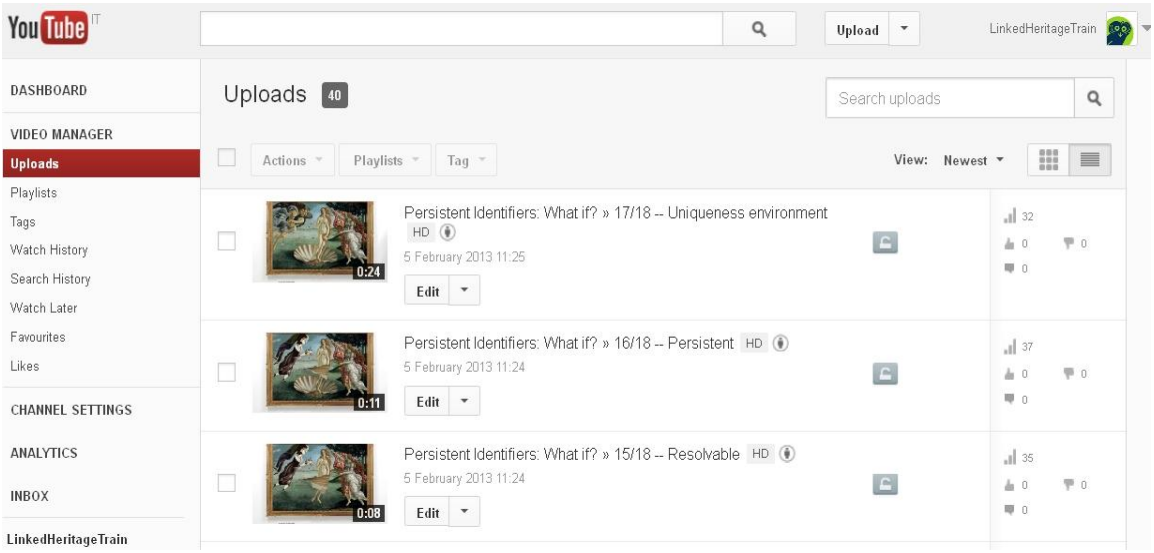


Figure 5 – YouTube hosts the Linked Heritage Learning Object embedded videos

The pages were created with clarity, portability and standards in mind, starting from the tools chosen to produce codes, images and videos. A particular effort was made to create pages compliant with website accessibility recommendations.

3.1.2 Learning objects produced

Production of Linked Heritage Learning objects: different types of skills

Setting up a learning object requires great effort from several points of view: content, resources, design and technical skills. In fact many activities, and associated expertises, are strictly necessary in order to create one learning object: concept, storyboard, graphic design, content draft, language revision, sound effects collation, music editing and recording, and video editing. Furthermore, legal issues must be carefully considered: before using any sound track or picture, it must always be checked they are not protected by any form of copyright.

The Linked Heritage learning objects were produced thanks to the support of music composers and performers, multimedia project experts, young geeks, photographers, camera operators, set designers and illustrators who gave the project their talents, availability and work for free.

Figure 6 – Experts who gave the project their talent and availability for free

UNIPD Linked Heritage team: collaboration experience

The University of Padova Linked Heritage team collaboration started in January 2012, at the beginning of the project. Until then, the University of Padova Library System did not have a team dedicated to the development of learning objects, even though Unipd librarians already had the necessary skills and expertise in multimedia and web-based training course production.

To improve the technical skills required to fulfil the project tasks, the team members attended a special workshop on video production with free software Stopmotion for animation and Cinelerra for video editing (Ubuntu Linux Operative System). During this workshop the team produced the promotional video "Why and how to participate to Europeana", currently available on You Tube[3], that was included in a Learning object with the same title[4].

The following picture sequence shows the main steps of the video production. They also show the UNIPD Linked Heritage team at work.

3 http://www.youtube.com/watch?v=CPaMZFW1hxQ&feature=player_embedded

4 http://linkedheritage.cab.unipd.it/training/LO-04/en/00.html

Figure 7 – UNIPD Linked Heritage team producing the Stopmotion animation "Why and how to participate to Europeana"

Learning object. Persistent Identifiers: What if?

Although this Learning object is aimed at managers of cultural institutions, it is also interesting for entry level LIS professionals.

Its contents, supplemented by training materials and further readings, raise awareness and provide users with basic information on Persistent identifiers (PIDs) and their applications in digital collections of cultural institutions.

This Learning object deals with PIDs. Texts are mainly based on the Athena project booklet *Persistent Identifiers (PIDs): recommendations for institutions* by Gordon McKenna and Roxanne Wyns[5] and present the main results of the Athena Project 2008-2011.

It is divided into 17 sub-units each of which consists of two parts: on the left of the screen, a dialogue between two owlets introducing PIDs and, on the right of the screen, a video showing the concepts.

We came up with the idea to use the famous painting by Botticelli, *The Birth of Venus*, as a metaphor. Each PID functional requirement is represented by a visual metaphor associated with a musical metaphor: uniqueness, persistency, resolvability, reliability, authoritativeness, flexibility, interoperability and cost effectiveness.

5 http://www.athenaeurope.org/getFile.php?id=779

The section “Readings and training materials”, inside the Virtual Learning Environment includes a bibliography and booklets by ATHENA WP3 Working Group and by Work package 2 of the Linked Heritage project.

Persistent Identifiers: What if? is currently available in the following languages: English, French, German, Greek, (Italian), Polish, (Sweden) and Spanish. Translation among brackets are ongoing at the date of this deliverable and will be ready by the end of the project.

Figure 8 – The cover of the Learning object “Persistent Identifiers: What if “

Learning object. Digitisation Life Cycle

This Learning object is both for managers of cultural institutions and technicians involved in digitisation projects.

Managers can have an overview of the entire workflow, while librarians, archivists, technicians are given practical examples of activities/situations they might face in the digitisation process which nowadays affects many cultural institutions.

Its content, supported by training materials and further readings, raises awareness and, at the same time, provides technical information on digitisation workflow of cultural institutions' collections.

This learning object presents the digitisation workflow both in theory and in practice. It consists of two parts: the first one gives an overview of the entire digitisation workflow based on the digitisation guidelines of the University of Padova Library System provided by the Phaidra working group. The second section focuses on a case study of the University of Padova the "Botanists portrait collection".

The case study is also illustrated by two videos showing the preservation of items and digitisation.

The section “Readings and training materials” inside the Virtual Learning Environment contains a bibliography and guidelines by ATHENA WP3 Working Group and by MINERVA-eC project.

*Digitisation Life Cycle i*s currently available in the following languages: English, German, Greek, Italian, Polish, (Swedish), Spanish. Translation among brackets are ongoing at the date of this deliverable and will be ready by the end of the project.

Figure 9 – The cover of the Digitisation Life Cycle Learning object

Learning object. MINT Services

It is aimed at technicians (librarians, archivists, museum curators) involved in metadata mapping procedures. It could be a useful resource also for managers of cultural institutions because they can gauge the skills required by their staff, and decide if they can contribute to Europeana independently or with the help of an aggregator.

The resource comprises 4 sections: Aggregators, the Linked Heritage aggregator, MINT[6] platform and Practical tips.

The Linked Heritage aggregation methodology and the mapping workflow in MINT - the technological platform developed by the National Technical University of Athens (NTUA) - is thoroughly described along the learning object. The user can follow step by step the mapping workflow by means of a sequence of screencasts realised by the NTUA team. Furthermore the learning object gives practical tips to technicians dealing with specific mapping activities as for example how to set metadata and digital object rights in MINT. Also in this case each mapping activity is supported by a detailed description and a

6 MINT (Metadata Interoperability) Services is the Linked Heritage technological platform: http://mint-projects.image.ntua.gr/linkedheritage/

screencast realised by the UNIPD Linked Heritage team. The learning object is enriched by the training materials and further readings.

The section “Readings and training materials”, inside the Virtual Learning Environment, includes a bibliography as well as guidelines and manuals by MINT services.

MINT Services is currently available in the following languages: English, Spanish and Swedish,

Figure 10 – The cover of the MINT Services Learning object

Learning object. Why and how to contribute to Europeana

It is specifically tailored for cultural institution decision makers. Why and how to contribute to Europeana describes the motivations for which cultural heritage institutions should contribute to Europeana. It also examines the University of Padova experience. The learning object contains the description of our Atheneum’s main technical steps, the workflow and some administrative information.

Why and how to contribute to Europeana is currently available in English and Sweden.

Figure 11 – The cover of the Why and How to contribute to Europeana Learning object

Learning object. Persistent Identifiers: Commercial and heritage views

It is a series of case studies on persistent identification presented by Linked Heritage partners led by EDItEUR. These studies point out the differences and similarities between the commercial media and cultural heritage sectors.

The intended audiences are both the library and information professionals & masters students and the cultural heritage (museum, archive, gallery) professionals & masters students.

Persistent Identifiers: Commercial and heritage views is currently available in English, Greek, (Swedish). Translation among brackets are ongoing at the date of this deliverable and will be ready by the end of the project.

Figure 12 – The cover of the Persistent Identifiers: Commercial and heritage views Learning object

Learning object. Terminology

This learning object introduces work package 3 of the Linked Heritage-project to unskilled or entry level learners in galleries, libraries, archives and museums.

Work package 3 (WP3) is dedicated to the exploitation of mono- or multilingual vocabularies in a digital environment. Thesauri are used as indexing tools, e.g., the subject index at the end of a book, and as inventory devices, allowing for structural and logical classifications of objects in a collection management system. They can be activated in a database or in an online catalogue as search tools enabling more detailed and accurate query results. Vocabularies can be implemented as additional browsing and navigation tools or as interactive term suggestion applications. They also facilitate the interaction of multiple databases. As the World Wide Web is gradually increasing its information content, specialists emphasise the importance of vocabularies as means to structure information on the web. The knowledge that is inherent to structured vocabularies can help retrieve information and search online, thus promoting the development of the semantic web or linked open data.

This tutorial explains the following subjects in a simple and practical way:

- the notion of terminologies and their importance in enhancing digital content
- step by step guidelines on how to improve your vocabulary on the world wide web
- an introduction to the terminology management platform (TMP)
- an introduction to the Simple Knowledge Organisation System (SKOS)
- an overview of the most recent literature and websites
- contact information

Figure 13 – The cover of the Terminology Learning object

Learning object. Public-Private Partnership with Europeana

This learning object is a very brief overview of the "state of play" for trade companies wishing to offer their product data to Europeana, with pointers to freely-available tools and documents contributed by EDItEUR to support this process as part of the Linked Heritage.

It describes:

- What Europeana is and does;
- Why it may be valuable to the commercial sector;
- Any relevant legal questions;
- Technical services available to support collaboration.

This learning object is currently available in Bulgarian, English, Spanish and (Swedish). Translation among brackets are ongoing at the date of this deliverable and will be ready by the end of the project.

Figure 14 – The cover of the Public-Private Partnership with Europeana Learning object

Learning object. Linking cultural heritage information

This Learning object focuses on the current structure of the Linked data Cloud - the best known representation of linked data. Size, licenses, subjects… the information was taken from the analysis made by Linking cultural heritage information, the 2nd Work package of Linked heritage.

Cultural heritage institutions dealing with linked data publication (museums, libraries, archives..) can find best practice recommendations.

Target audience:

- Decision makers and project managers of public libraries, museums, archives.
- Culture and information specialists from museums, libraries and archives.

Thanks to Linked data – especially Linked Open Data – they can enrich knowledge and improve their visibility on the web.

This Learning object is currently available only in English.

Figure 15 – The cover of the Linking cultural heritage information Learning object

3.1.3 Learning objects usability and accessibility: evaluation tests and enhancements

Several evaluation tests were set up to improve usability and accessibility of the Linked Heritage learning objects.

Between September 2012 and 6 March 2013, three sessions of evaluation tests have been organised:

- September 2012: a first session (observational method: test monitoring) was carried out in Padova by 6 librarians on the already implemented learning objects ("Persistent Identifiers: What if?"; "Digitisation Life Cycle");
- 7-25 January 2013: a second session was extended to the Linked Heritage e-learning Working Group by means of a dedicated questionnaire. The evaluation test was presented at the WP7 e-learning Working Group meeting held in Lisbon on 30 November 2012. A Virtual Learning Environment (VLE) "business card" with the URL of the Linked Heritage course in Moodle was handed out to each participant together with a personal account to log on to the Virtual Learning Environment. Participants performed the evaluation test by themselves connecting remotely. Thirteen questionnaires were filled in;
- 6 March 2013: a third session took place at the Emeroteca Ca' Borin of the University of Padova (6th March 2013). There were two groups of users: 1) librarians, museum curators and university students attending the Italian Workshop on Linked Heritage and 2) Linked Heritage partners attending the project Training Workshop on Learning Objects.

The session focused on the following learning objects: "Why and how to participate in Europeana", "MINT Services", "Digitisation Life Cycle" and "Persistent identifiers: What If?". Each participant was invited to explore the available tools (e-learning platform, learning objects and related contents) and

to fill in the evaluation questionnaires. Fifty-nine questionnaires were submitted and are currently preserved at CAB.

All the comments and suggestions received by the participants were analysed and, considered the project timeline, prioritised.

Learning objects **quality enhancements** were focused on the following issues:

Clearer and more comprehensive navigation

A global homepage (http://linkedheritage.cab.unipd.it/training/LO-00/en/overview.html) was created to facilitate navigation between different Learning Objects and from Learning Objects to the Virtual Learning Environment (Moodle). Although it displays a slightly different layout from the one used in Learning Objects, it is still recognisable.

This page explicitly states and briefly explains the separation between the Virtual Learning Environment and the group of Learning Objects. A short description is given for each available Learning object, alongside with the links associated with each translation and with the pertinent pages in the Virtual Learning Environment.

Each Learning object contains a navigational summary at the top of long pages, and proper icons show the links to external websites.

Accessibility

To ensure more accessible navigation, a hidden link was added to enable visually impaired users to skip past the sometimes long list of navigational links displayed at the top of each page.

The LANG attribute declaring the language of the page (i.e. lang="en" xml:lang="en") was added to the HTML tag.

Online tools to test accessibility and a live test performed by a colleague who uses a screenreader to browse the Web confirmed that the Linked Heritage Learning Objects site has achieved good accessibility level, at least with the browsers Explorer 8, Mozilla Firefox and Chrome.

Other

The misleading "Download transcript" link was removed from Learning object 1 "Persistent identifiers: What if?" because some users expected the videos to include speech, which is not the case.

In Learning object 1 "Persistent identifiers: What if?" all quotes from Gordon McKenna and Roxanne Wyns booklets are now linked to the associated PDF.

3.1.4 Translation into other languages

Some e-learning Working Group members agreed to translate one or more Learning objects from English into the following languages: Bulgarian, French, German, Greek, Italian, Latvian, Polish, Spanish, Swedish.

The translation process was quite simple. Partners were asked to carry out the following activities:

- Decompress the Learning Object ZIP file
- Duplicate the sub-folder (“en”) and rename it "lang"
- Translate all the HTML pages inside the newly created "lang" folder. It is highly recommended to use an editor that doesn't add any extra codes

- Translate all the files inside the sub-folder "text" (transcript of HTML pages and videos; SRT files for subtitles in YouTube)

- Send UNIPD back the zipped folder with the translated Web pages and YouTube subtitles

The following synoptic table summarises the Learning objects developed during the project and the partner organisations that contributed to the content editing and/or translation. At the date of this deliverables partners from Sweden and Latvia are finalizing their Learning objects translations. The results will be available soon.

Title	Content contributors to the English version	Translation contributors
Pids: What if	UNIPD	es (i2CAT), el (HMC), de (SPK), fr (Dedale), it (UNIPD), pl (ICIMSS), sv (RA)
Digitisation Life Cycle	UNIPD	es (i2CAT), el (HMC), de (SPK), it (UNIPD), pl (ICIMSS), sv (RA)
MINT Services	UNIPD, NTUA, ICCU	*es (i2CAT), lv (KIS), sv (RA)*
Why & how to contribute to Europeana	UNIPD, ICCU	sv (RA)
Pids: commercial & heritage views	EDItEUR with mEDRA, MVB, NSL, TIB	el (HMC), sv (RA)
Public-Private Partnership with Europeana	EDItEUR with mEDRA, MVB, NSL, TIB	bg (CL-BAS), es (i2CAT), *sv (RA)*
Terminology	KMKG, MCC	*sv (RA)*
Linked Data	UNIPD, CT	

Table 1: Partners involved in the Learning objects' implementation

The collaboration with all LH partners and in particular among e-learning working group participants has been extremely beneficial and brought many positive outcomes, as already underlined by EU Commission during the first LH Project review held in Brussels on 28th Sept 2012.

Among the main benefits and positive outcomes of collaboration among e-learning working group, we would like to stress the following:

- the active contribution in the definition of the Table of Contents of the Linked Heritage training syllabus granted by a collaborative network of experts and stakeholders;
- the strong effort put in the Learning objects translation activities which permitted the availability of the Learning objects in different languages, facilitating a wider dissemination of the Linked Heritage training programme to the potentially interested audience;
- the strong contribution in the Learning objects content creation and definition of the relevant training materials given by Linked Heritage professionals by suggesting bibliography, case studies and videos.

In conclusion, the statement of the Description of Work (p. 104), that is: "WG structure is flexible and supports open discussion and exploration, beyond the more rigid contractual nature of a work-package. It tends to encourage longer-duration, deeper discussion, often involving those relatively few individuals with a deep interest and expertise in a detailed topic. As result, only those who need to be involved in a particular discussion actually take part, ensuring both a high level of shared expertise and an efficient deployment of resources for all partners" is fully confirmed in the e-learning working group collaboration.

3.1.5 Reuse

The Linked Heritage Learning objects are open, web-based educational resources that can be delivered not only through a Virtual Learning Environment but also through multimedia Web channels.

Openness and re-usability of these Learning objects were therefore considered essential requirements in order to allow other course creators and teachers to **reuse, modify and/or translate** them and maximise impact on the intended audience.

A link to **instructions** on how to edit and/or translate the project Learning objects was therefore included both in the Main page of the Linked Heritage course in Moodle (Cf. D7.4, §2.5.3) and in the Linked Heritage training programme homepage (Cf. §3.5).

Moreover, to prompt reuse of Learning objects, the LH course was made available in the **FreeLoms Repository**, a collaborative environment for sharing and creating Learning Objects in an open perspective designed within the Sloop project[7] (2005-2007) - Sharing Learning Object In An Open Perspective - a project co-funded by European Commission under the Leonardo Da Vinci programme.

7 http://www.sloopproject.eu/sloop/

Figure 16 - Instructions on how to modify the Linked Heritage Learning objects

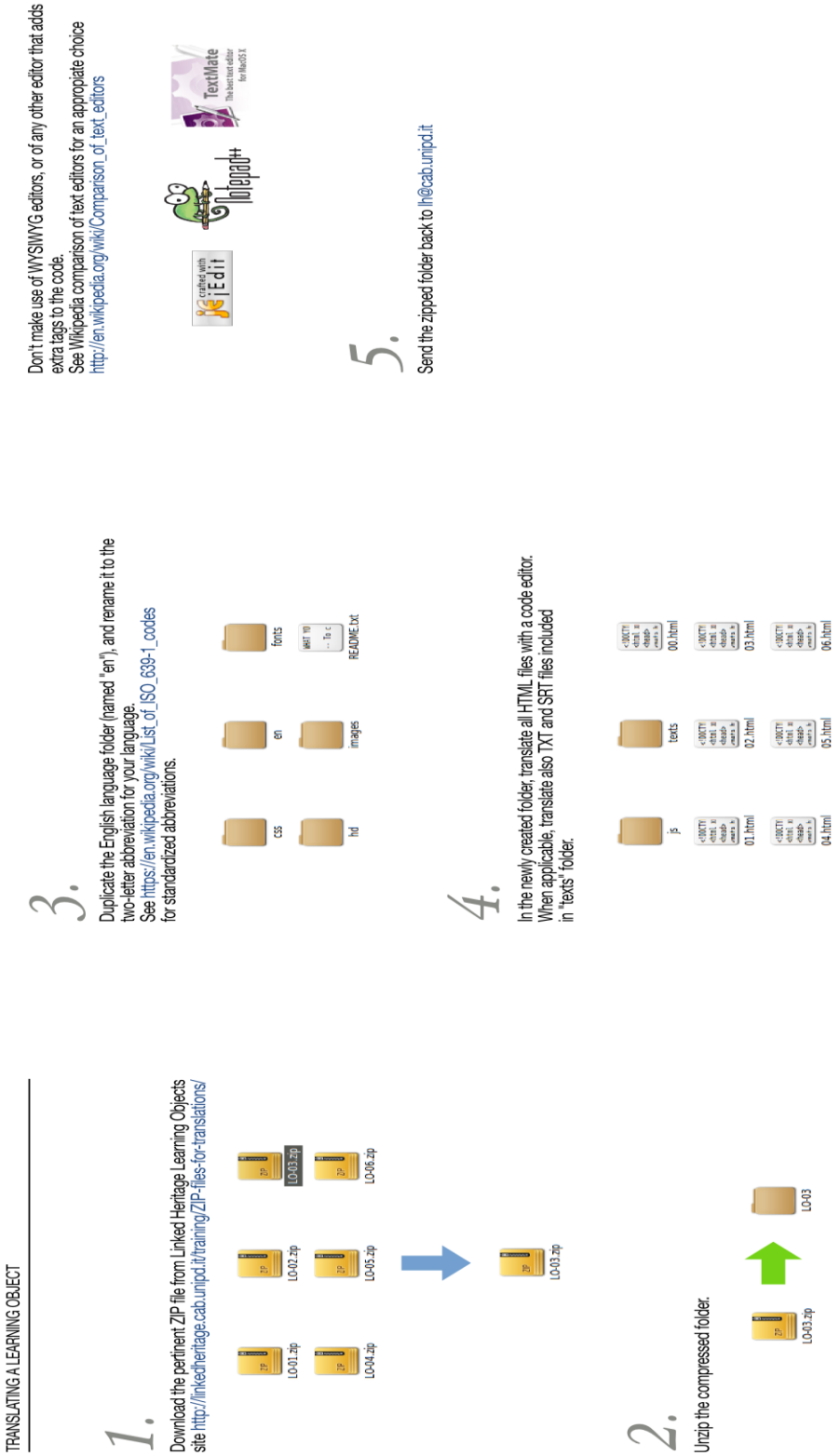


Figure 17 - Instructions on how to translate the Linked Heritage Learning objects

3.1.6 Assessment and Evaluation

Assessment processes improve teaching and learning, and this is all the truer for online learning than for one-to-one, in which assessment is often viewed as a separate task from learning. The principles of assessment have developed in one-to-one environments and have not changed in online environments. However, the differences between these modes challenge the implementation of assessment practices. Traditional assessment places learners as recipients of knowledge, where learning is measured and documented at the first level of Bloom's taxonomy as knowledge and comprehension. The role of assessment activity in online learning can help improve the e-learning process for all participants. It represents a feedback source for learners (giving scores and feedback), for instructors (by giving support and feedback on teaching) and for instructional designers (to update the contents of the learning system, to adjust and develop technical aspects). The assessment activity as an integrating step is supported by the monitoring opportunities that provide a global view of scores and personal steps by individual interactions with the system.

The evaluation process integrates the information supplied by the assessment activity by requiring user s to think about their perception of the online learning experience taking place. Thus, assessment and evaluation work together in developing the learning system.

In our case, the level of learners' interaction is basic, as the active role of learning does not lie so much on peer or system interaction, but on the personal management of learning, reflection, time management, and self-assessment. This is consistent with an assessment modality based on knowledge and comprehension. Future implementation should witness further steps towards greater interaction with the system and peers associated with a wider range of assessment tools.

In Linked Heritage course each Learning object comes with two questionnaires:

Assessment: it establishes how much users have learned from the proposed Learning object.

Assessment of the Learning object was designed to both give a learning self-assessment tool and stimulate the revision of concepts dealt with in the Learning object. The concepts covered in the Learning objects were identified, re-examined and subjected to questions, often with multiple choice answers. The same concepts were then subjected to further True/False questions to reinforce their acquisition. We felt that a successful outcome of the assessment required careful analysis of the Learning object, thus making it a fundamental accompanying tool to maximise learning results.

Evaluation: it establishes what users think of the Learning object and aims to evaluate the work done by the Learning object implementers.

The Learning object evaluation questionnaire aims at gathering users' feedback on technical and conceptual aspects of the learning objects through questions on the adequacy and clarity of the content, the appropriateness / adequacy of the accompanying documentation, the extent to which technical issues support the efficacy of the Learning object like audio, fonts, colors, subtitles, video editing aspects, navigation tools and interaction.

An example of the assessment and the evaluation questionnaire is available in D7.4 Annex1.

3.2 TRAINING MATERIALS STRICTO SENSU

These materials are essential educational resources as they improve knowledge and understanding of one or more specific subjects.

They are:

1) **public deliverables and booklets** created within the Linked Heritage project;
2) **guidelines, manuals edited by Linked Heritage WP Leaders** and other subject experts - both in English and in their native languages - that they consider important reference materials to be included in the Linked Heritage course in Moodle;
3) **guidelines, manuals and booklets edited by EU funded sister projects** e.g. Athena and Minerva.

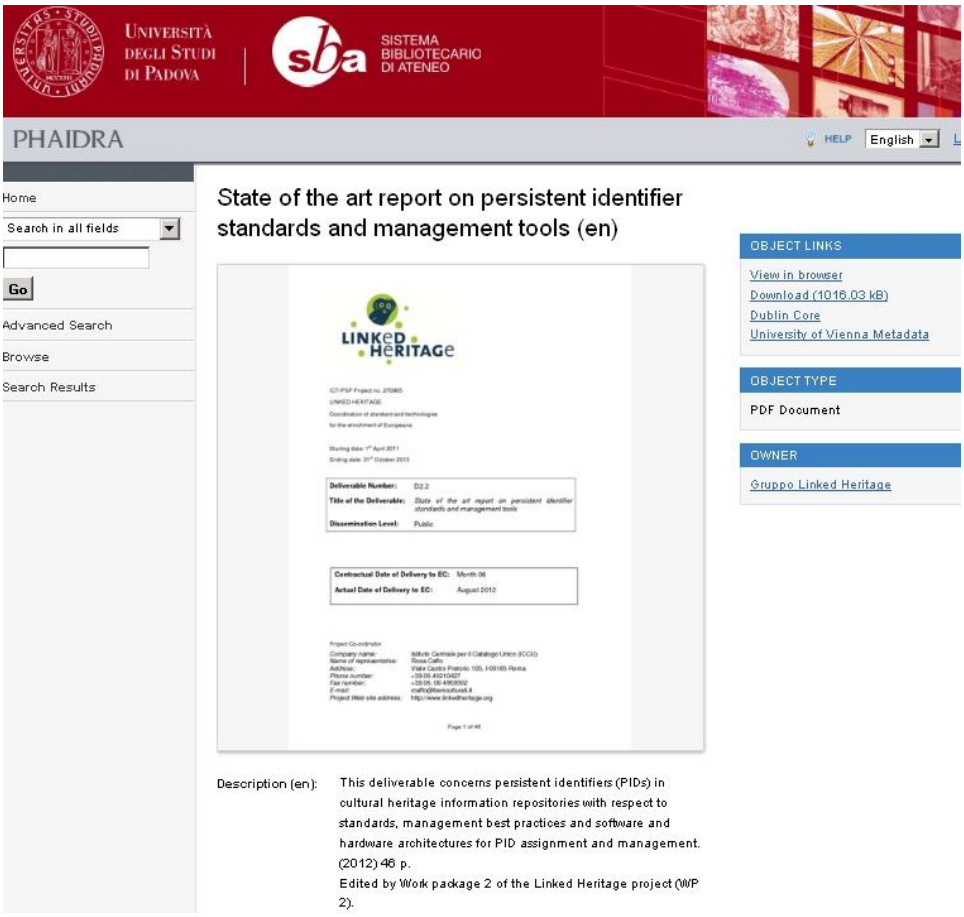


Figure 18 - Example of public deliverable created within the Linked Heritage project

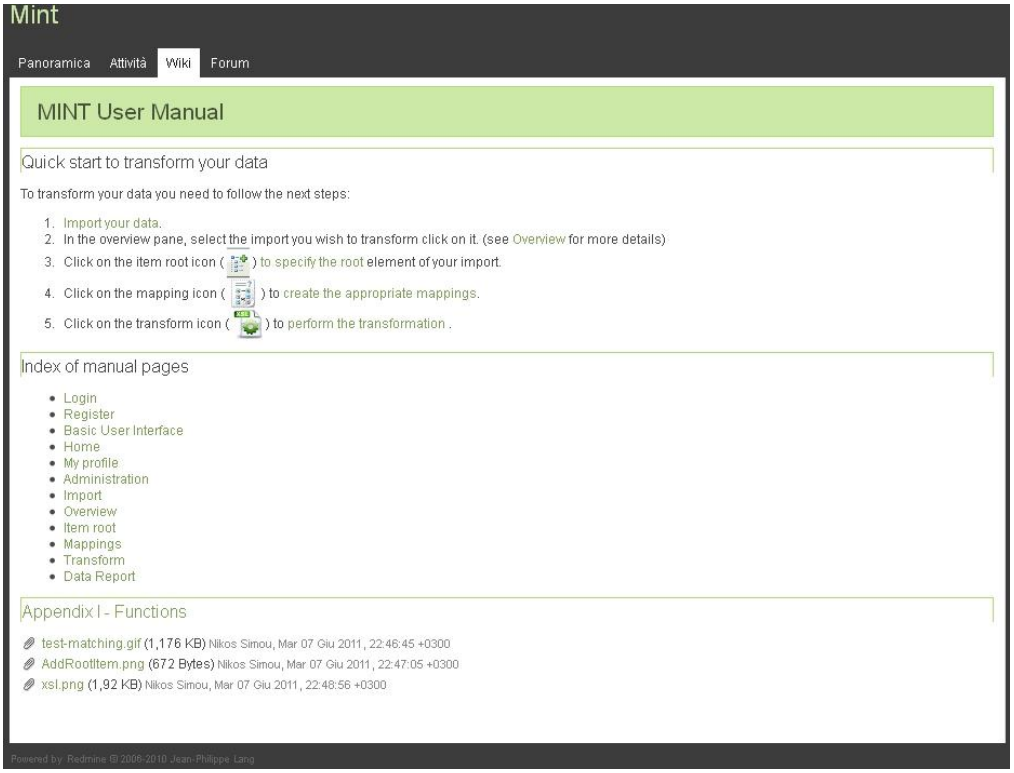


Figure 19 - Example of manual edited by Linked Heritage WP5 Leader

Figure 20 - Example of booklet edited by sister projects (Athena)

3.3 CASE STUDIES

In addition, we felt we had to include documentation (articles published in journals, technical reports, papers and posters in conference proceedings) offering instructive examples of digitisation projects, persistent identifiers systems adoption, Linked Data production, metadata mapping, etc., as a useful resource to improve the learners' know-how and skills.


JLIS.it
Italian Journal of Library and Information Science
Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione

HOME ABOUT REGISTER CATEGORIES SEARCH CURRENT ARCHIVES
ANNOUNCEMENTS SUBSCRIPTIONS THESIS ABSTRACTS

Home > Vol 4, No 1 (2013) > **Agnoloni**

Publishing Value Vocabularies and Standard as Linked Data

Annotation schema for legal doctrine: a case study on DoGi database

Tommaso Agnoloni, Ginevra Peruginelli, Maria Teresa Sagri, Daniela Tiscornia

Abstract

Interoperability today is the key term for the enhancement of databases published on the web: the data, when isolated, have little value, on the contrary, their value increases significantly when different datasets, produced and published independently by different providers, can be reused and freely mashed by third parties. The use of data for new purposes not foreseen by organizations and individuals who publish "raw data" is one of the advantages of linked open data model. To achieve these benefits content and relationships between the entities described in the dataset should be explicitly represented in standard web formats (XML, RDF, URI). The DoGi-Legal Literature database, one of the most valuable sources for online access to legal doctrine, created and managed by the Institute of Legal Information Theory and Techniques of the CNR is following this direction. This paper will define the schema of the data representing the database in RDF format. This will make the DoGi database interoperable between different data and service providers (libraries, publishers, information services for accessing national and European legal information), allowing the creation of new advanced services to be made available on the web of data. In particular, the contribution will focus on the goal to promote semantic interoperability between the DoGi classification schema and other semantic indexing tools in legal domain.

Figure 17 - Example of training material: article dealing with a case study

3.4 BIBLIOGRAPHY

This list of selected works offers learners in-depth analysis of specific topics. In the Linked Heritage Virtual Learning Environment, each topic has its own bibliography, which includes sources used for Learning object content editing and works selected by Linked Heritage specialists.

The Linked Heritage Training Programme references comprise:

- Europeana Documentation
- European Projects booklets
- lists of documents provided by experts on the topics of the project
- papers and posters published in recent Conference Proceedings.

In addition, bibliographic search sessions were carried out in:

- the most relevant UNIPD academic databases as Scopus, Web of Science, Compendex
- relevant scientific search engines, like Google Scholar, Base, Scirus
- subject repositories like E-LIS, OAIster (OCLC)

other sources for bibliography were:

- Europeana ecosystem Projects (e.g. Europeana for Education)
- conferences organised by Linked Heritage projects and sister projects

The Linked Heritage Training programme Bibliography was drawn up using the Harvard referencing style. References are listed alphabetically by the author's surname. Whenever possible, full text URLs were added to the citation.

Formats and software sources are placed at the beginning of the bibliography to facilitate their visualisation.

The bibliography was uploaded and managed by the UNIPD Linked Heritage team with the help of RefWorks, an online reference management software package, and then exported in the Linked Heritage course on Moodle, where they it can be retrieved by the learners through the Linked Heritage course Google Custom Search engine (Cf. D7.4. §2.5.5)

Figure 22 - screenshot of the bibliography in the Moodle Linked Heritage Course

3.5 THE HOME PAGE OF THE LINKED HERITAGE TRAINING PROGRAMME AND LEARNING OBJECTS

We created the Homepage of the Learning objects[8] and Linked Heritage training programme to facilitate navigation between the Learning package in Moodle and the individual Learning objects available online. This Webpage is both a gateway between the Virtual Learning Environment and the Learning objects, and a dissemination tool to which any partner institution can link from their webpages.

8 http://linkedheritage.cab.unipd.it/training/LO-00/en/overview.html

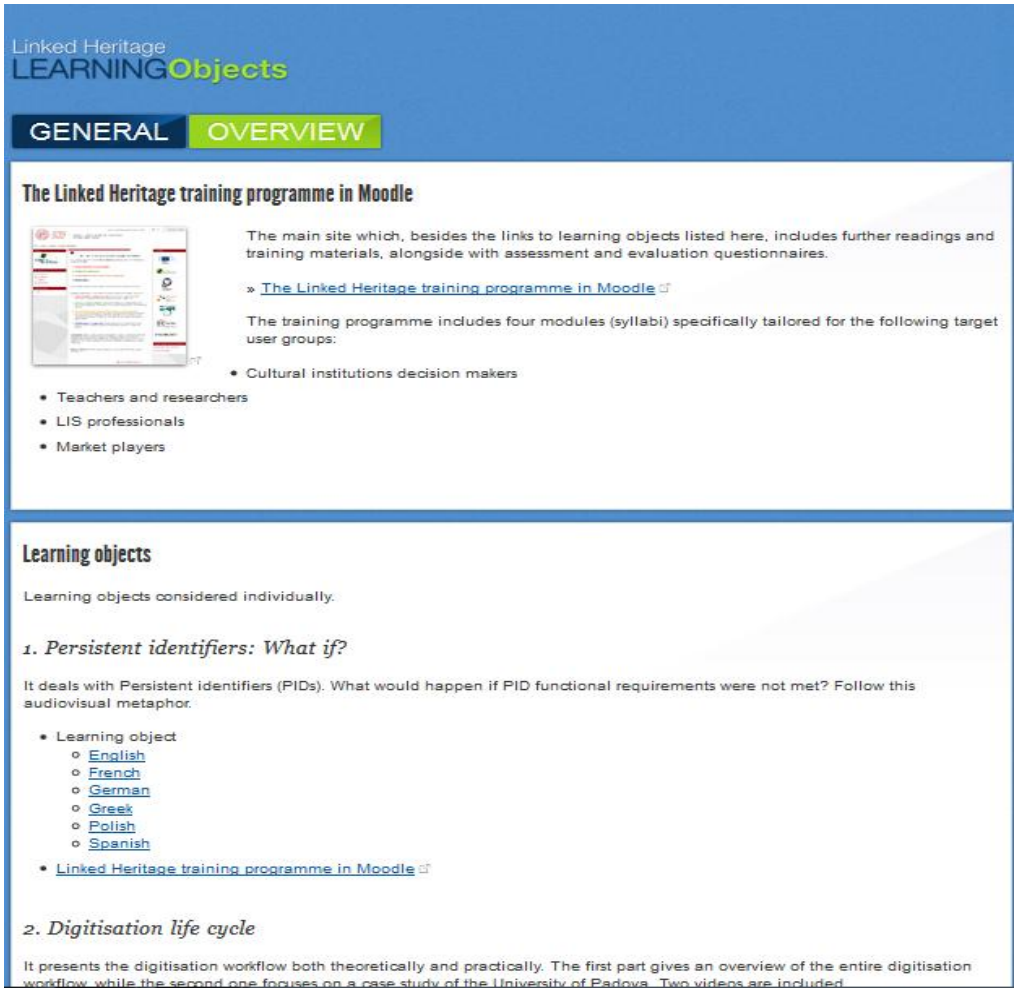


Figure 24 - the Homepage of the Linked Heritage training programme and Learning Objects

3.6 EXPECTED LEARNING OUTCOMES

Learning outcomes on successful completion of the Linked Heritage programme:

- **cultural institution managers and decision-makers** can appreciate what Europeana is and therefore decide whether they want to apply;

- **teachers, educators and researchers** can gain knowledge and understanding of the scope of Europeana and of its content, how to search and exploit it for educational purposes;

- **Library and Information Science (LIS) students, professionals and technicians** can appreciate the importance of professional development and the educational resources available online and selected for this Syllabus by the Linked Heritage experts, to keep up to date with the most advanced issues on digital libraries.

- **market players** can appreciate what Europeana is and does, and the main results of the Linked Heritage, in particular the technical support for trade partners.

3.7 PRESERVATION IN PHAIDRA

In addition, the Learning objects and training materials senso strictu will also be hosted and downloadable (ZIP files for Los and PDF files for documents with text and images) in **PHAIDRA**[9] **(**please, see Linked Heritage deliverable D7.4, figure 10**).**

Phaidra is the acronym for **Permanent hosting, archiving and indexing of digital resources and assets**. It is a digital asset management system with long-term archiving functions.

Describing and indexing digital objects in Phaidra is simple and fast. Each resource in the database can be easily cited thanks to a permanent link.

The platform is based on the Fedora Commons Repository software. The formats adopted for the metadata and the digital objects guarantee long-term data archiving and interoperability with other systems, i.e. Europeana, as requested by DOW. Security is guaranteed both by the well-defined terms of use and the licensing system that regulates the upload of digital objects.

3.7.1 Why the University of Padova adopted Phaidra

Several digitisation projects carried out by the University of Padova Library System emphasised the need to find a comprehensive system:

- to manage a long-term digital assets archiving service;
- to enhance the University of Padova past and present intellectual production not subjected to private economic interest;
- to manage wide collections of multimedia objects through an access control system

Hence, the Phaidra project by University of Vienna was selected and a profitable collaboration with the University of Padova started in 2010[10].

3.7.2 Who can access Phaidra?

Anyone can search and browse archived digital objects, while uploading and management of different digital objects (images, documents, books, audios, videos, collections, teaching resources) requires authorisation. Researchers from other universities and research institutions can upload their contributions to Phaidra through guest accounts.

In addition, the Phaidra Importer application makes the publication of online books, leaflets, and pamphlets easier.

All the digital objects catalogued in Phaidra are provided with a permanent identifier and a rich set of metadata (Learning Object Model format). In Phaidra we created a collection called "Linked Heritage: booklets and other resources" to gather all the Learning objects and training materials senso strictu that were created or selected for the Linked Heritage Training Programme.

9 http://phaidra.cab.unipd.it

10 http://www.phaidra.org/

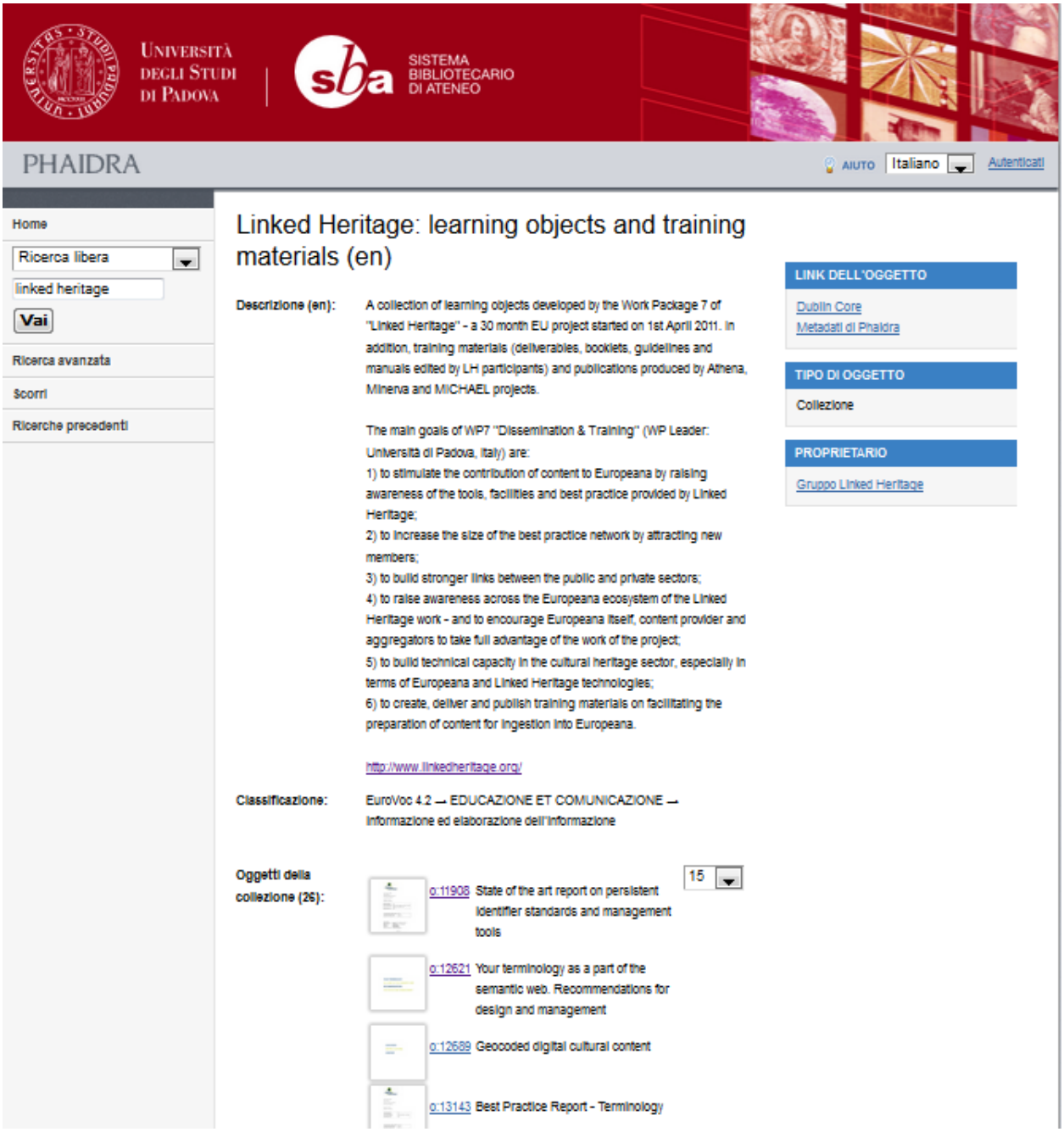


Figure 25 – Screenshot of the Linked Heritage Training Material collection in Phaidra.

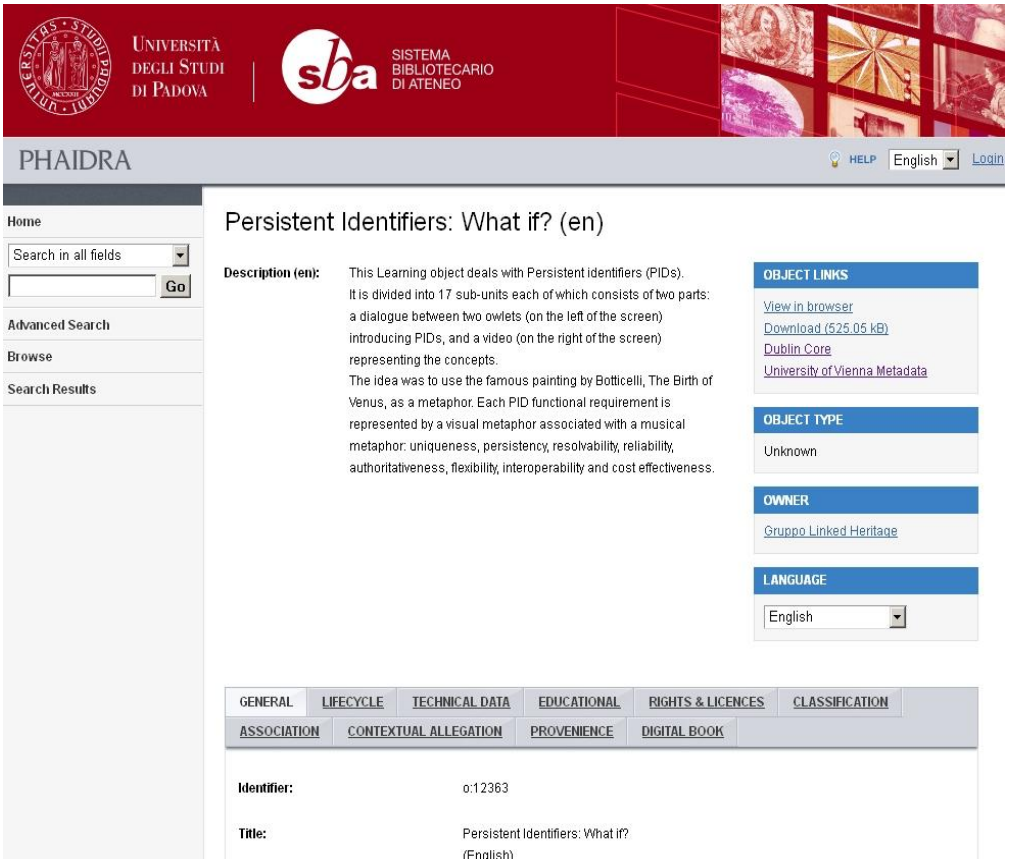


Figure 26 – Example of a Linked Heritage Learning Object description in Phaidra.

3.8 IPR

The booklets, public deliverables, learning objects and training materials created within the project will be published under a CC-BY-NC-SA license.

In line with the EC open source policy initiatives, all the software programmes used for the production of the Learning objects are Free Software/Open Source (GNU Project and/or OSI definition):

Blender

Cross-platform 3D creation software providing a broad spectrum of modelling, texturing, lighting, animation and video post-processing functionality.

Cinelerra CV

An advanced non-linear video editor and compositor for Linux.

Gimp

Cross-platform GNU Image Manipulation Program. A piece of software for such tasks as photo retouching, image composition and image authoring.

Inkscape

Cross-platform vector graphics editor using the W3C standard Scalable Vector Graphics (SVG) file format.

jEdit

Cross-platform mature programmer's text editor customizable with plugins.

Linux Stopmotion

A program for creating stop motion animation movies either from hard drive pictures or from webcamera pictures.

Moodle

PHP web application for producing modular internet-based courses that support a modern social constructionist pedagogy.

Sozi

A small program that can play animated presentations. The presentation editor is an extension for the Inkscape vector graphics editor.

xvidcap

A small tool to capture things going on on an X-Windows display.

4 ACCESS POINTS AND TRAINING MATERIAL PROMOTIONAL ACTIVITIES

4.1 ACCESS POINTS

At the end of the project users will be able to **access** the Learning objects and associated training materials senso strictu in many ways, and therefore boost the dissemination of Linked Heritage educational resources and improve their retrieval in the Web. The access points provided so far are:

- the Linked Heritage Website;
- the University of Padova and Linked Heritage partner institutions Websites and multimedia channels
- the Virtual Learning Environment of the University of Padova;
- the University of Padova PHAIDRA platform;
- YouTube;
- FreeLOMS, the Learning Object repository developed within the EU-funded projects SLOOP and SLOOP2DESC.

4.2 PROMOTIONAL MATERIALS

During the project, the University of Padova Linked Heritage team came up with the idea to produce, in addition to the Learning Objects, two multimedia resources devoted to the dissemination of the project results.

One is a video presenting the Linked Heritage Virtual Learning Environment and the first two Learning Objects created: “PIDs:What if?” and “Digitisation Life Cycle” [11]The video was shown as a backdrop at the Italian Workshop on Europeana and CulturaItalia held in Padova on 6th March 2013 and other local events[12].

The other is a Stopmotion animation promoting Europeana[13] which was also included in the Linked Heritage Learning Object “Why and How to contribute to Europeana”.

In addition, a set of postcards[14] (digital and printed) – one for each Learning object – were produced to promote the eight Learning objects produced by the Linked Heritage project,

11 http://www.youtube.com/watch?y=AcHlJe8Vt-k

12 http://www.linkedheritage.org/getFile.php?id=432

13 http://www.youtube.com/watch?v=CPaMZFW1hxQ&feature=player_embedded

14 http://www.linkedheritage.org/index.php?en/211/learning-objects

5 CONCLUDING REMARKS

The WP7 Task 7.2 Training has closely liaised with the WP7 Task 7.1 in the last phase of the project (15 July - 30 September 2013) to elicit follow ups between partners and external communities.

We think that the Linked Heritage Training Programme, also thanks to the strong **dissemination activity**, will successfully

- raise awareness among cultural institution decision makers on Linked Heritage main results and how to participate in Europeana and share best practices,
- inform and promote Europeana among teachers
- introduce entry-level users to hot topics from cutting-edge digital libraries,
- provide authoritative reference sources and contents to library, museum and archive specialists, validated by Linked Heritage experts.
- promote a mutual understanding between public and trade sectors on metadata, persistent identifiers standards, and best practices.

In addition to dissemination, the Linked Heritage Project focused on an **innovative task** aimed at the **e-learning training of users interested in digital cultural heritage themes** as well as in providing content to Europeana - mainly the content-contributing community, the government and policy bodies, the private (publishing) sector and, to a certain degree, the research community.

Since the start of the project (April 2011), the objective of the University of Padova has been to create a programme with top-quality content for self-training within an open-access, easily accessible and usable Virtual Learning Environment for an international audience of professionals, policy makers and Cultural Heritage experts.

In particular, this training programme aims at teaching users who are skilled in some Cultural Heritage areas (e.g. museum curators) but need to learn how to transfer their materials from a physical to a digital environment.

The positive feedback we got from archives, libraries and museum operators during the Italian Workshop "The digital cultural heritage towards Europeana and CulturaItalia: technical aspects and methods", which also included a Linked Heritage VLE and LO workshop, shows how the LH project satisfies a deeply felt need.

Transfer of knowledge between areas of the digital cultural heritage sector and improved competences will therefore breathe new life and bring a growing number of contents to Europeana.

By participating in the Linked Heritage Project, the CAB could discuss with the University of Vienna - its technological partner in the Phaidra project – the issue of **preservation and visualisation of multimedia objects** and the model to be used for multimedia resources (such as, but not limited to, Learning Objects) in Phaidra, the platform for the preservation of digital objects that the two universities share together with the University of Belgrade and 5 other Western Balkans universities.

The LO files of LH are arranged in folder structure and are currently preserved in Phaidra as zipped files.

Our work with our Austrian colleagues will produce a new detailed description of the characteristics of multimedia objects (e.g. structure, metadata, PID, relations, etc.), and an analysis of both suitable viewers and of the zipped file formats suitable for preservation.

We believe that the online “EU project: Linked Heritage” course, which was developed by the UNIPD with the support of the LH partner consortium, could be considered as an example of OpenCourseWare[15], precursor of MOOC (Massive Open ONline Course), the extremely interesting initiative started by Stanford University in autumn 2011 and that has now reached Europe with the first Pan-European University proposal with the collaboration of 11 partners and the support of the European Commission (www.openuped.eu).

In the **future**, MOOCs will play an important role in self-learning in the professional world. In particular, constant updating will be indispensable for operators and experts in the digital library sector.

We believe the training programme tailored to Linked Heritage - open access to a VLE with selected top-quality training materials that include best practices, methodological aspects and reference to standards - may be used as a model to tackle the next important challenges like the management of “big data” and 3D digitisation.

15 http://en.wikipedia.org/wiki/OpenCourseWare

6 BIBLIOGRAPHY

This is the comprehensive bibliography of the D7.3 "Training materials" and the D7.4 "Virtual Learning Environment" deliverables.

Akeroyd, A. 2005, "Information management and e-learning: some perspectives", *Aslib Proceedings: New Information Perspectives*, Vol. 57, No. 2, pp.157-167.

Anderson, T. & Elloumi, F. (eds.) 2004, *The Theory and Practice of Online Learning*, Athabasca University Press.

Blaschke, L. M. 2012, "Heutagogy and Lifelong Learning: A Review of Heutagogical Practice and Self-Determined Learning", The International Review of Research in Open and Distance Learning, Vol.13, No.1, pp. 56-71. Available at:

http://www.irrodl.org/index.php/irrodl/article/view/1076/2087 [accessed 26/09/2013]

European Union: European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Rethinking Education: Investing in skills for better socio-economic outcomes*, 20 November 2012, COM(2012) 669 final. Available at:

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0669:FIN:EN:PDF

[accessed 26/09/2013]

Council of the European Union, Council conclusions on investing in education and training — a response to *'Rethinking Education: Investing in skills for better socio-economic outcomes' and the '2013 Annual Growth Survey' 2013*, OJ C 64 5.3.2013. Available at:

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2013:064:0005:0008:EN:PDF [accessed 26/09/2013]

Council of the European Union, *Council conclusions of 26 November 2012 on education and training in Europe 2020 — the contribution of education and training to economic recovery, growth and jobs*, OJ C 393 19.12.2012. Available at:

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:393:0005:0007:EN:PDF [accessed 26/09/2013]

European Commission 2008, *Explaining the European Qualifications Framework for Lifelong Learning.* Available at:

http://ec.europa.eu/education/lifelong-learning-policy/doc/eqf/brochexp_en.pdf

[accessed 26/09/2013]

European Commission 2006, *Information Society and Education: Linking European Policies*, Luxembourg: Off. for Official Publ. of the Europ. Communities, 15 p.

European Union: European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Rethinking Education: Investing in skills for better socio-economic outcomes*, 20 November 2012, COM(2012) 669 final. Available at:

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0669:FIN:EN:PDF [accessed 26/09/2013]

Fidishun, D. (no date, circa 2005), *Andragogy and technology: Integrating adult learning theory as we teach with technology*, Malvern, PA: Penn State Great Valley School of Graduate Professional Studies.

Ravotto, P. 2012, "Dal corso eLearning alla comunità di pratiche. Esperienze di formazione insegnanti con Moodle in progetti europei", *Brick*, Vol. 1, pp.106-115.

Web Content Accessibility Guidelines WCAG 2.0, World Wide Web Consortium (W3C) recommendation, 11 December 2008. Available at:
http://www.w3.org/TR/WCAG20 [accessed 26/09/2013]

ANNEX 1: LIST OF VIDEOS IN THE LINKED HERITAGE LEARNING OBJECTS

PERSISTENT IDENTIFIERS: WHAT IF?

Introduction http://www.youtube.com/watch?feature=player_embedded&v=2RgQpWPfFik

What if: Global uniqueness http://www.youtube.com/watch?feature=player_embedded&v=F_58boTFDK4

What if: Persistence http://www.youtube.com/watch?feature=player_embedded&v=xZUwMIfjPis

What if: Resolvability http://www.youtube.com/watch?feature=player_embedded&v=ODlw137fZ_4

What if: Reliability http://www.youtube.com/watch?feature=player_embedded&v=6oY7JOICkzU

What if: Authoritativeness http://www.youtube.com/watch?feature=player_embedded&v=1bUUjcA-HUg

What if: Flexibility http://www.youtube.com/watch?feature=player_embedded&v=uFL2NtkjdPA

What if: Interoperability http://www.youtube.com/watch?feature=player_embedded&v=RHvXfONthfE

What if: Cost effectiveness http://www.youtube.com/watch?feature=player_embedded&v=rDTUL-FZ6Nw

Cost effective http://www.youtube.com/watch?feature=player_embedded&v=4SFjRyBSz5w

Interoperable http://www.youtube.com/watch?feature=player_embedded&v=Joha79WtekQ

Flexible http://www.youtube.com/watch?feature=player_embedded&v=8AKUDn8eonA

Authoritative http://www.youtube.com/watch?feature=player_embedded&v=XQ5fEXskqP0

Reliable http://www.youtube.com/watch?feature=player_embedded&v=0L0QvcsaRYs

Resolvable http://www.youtube.com/watch?feature=player_embedded&v=SEEhucX5R3w

Persistent http://www.youtube.com/watch?feature=player_embedded&v=P6ygaThSbuY

Uniqueness environment http://www.youtube.com/watch?feature=player_embedded&v=JNBFd3_tipM

DIGITISATION LIFE CYCLE

Preservation of items http://www.youtube.com/watch?v=2iTYI7TxwpE&feature=player_embedded

Video transcription (English translation):

==Housing and photographic restoration of the Botanists portrait collection==

===Housing===

The item is removed carefully from the original envelope. First, any loose dust is blown away using a rubber bulb syringe, then both sides are cleaned with a soft bristle brush, applying no pressure whatsoever to the surface. The location is marked in pencil on the original, and the item is then placed in a preservation-quality envelope.

===Restoration of glass plate negatives====

First, the plate is observed under a horizontal beam of light, on both sides; the characteristics of the item â€" and any damage â€" are recorded on a special chart. In this instance, inspection reveals the presence of silver mirroring and traces of graphite.

The glass side is cleaned with a solution of surfactants in water to remove surface dirt (such as dust and grease).

The emulsion side is cleaned with solvent, which can help to reduce the visual impact of the silver mirroring at least in part.

The plate is dusted off with an antistatic brush, housed in a suitable four-fold enclosure, and placed in a preservation-quality box.

In this instance the plate is broken and has been patched up with adhesive tape to keep the pieces together. The adhesive has degraded, and the tape must first be peeled away dry from both sides.

The next step is to take a strip of adhesive tape with a backing material and a type of glue suitable for contact with photographic material, and prepare a number of small transparent sticking plasters that will not show up in prints; these are applied to the glass side only, and will hold the fragments together.

===Restoration of positives===

Here are some positive photographic prints that have been mounted in frames.

The problems created by poor preservation are immediately clear from the type of mounting: not only is the material unsuitable, but there is a lack of uniformity in the supports utilized.

All the photographs restored were in general dirty, with clearly visible traces of dust, insect droppings, and in some cases larvae, which were also found on the backing materials.

First, the dirt is removed from all of the materials; the prints, in particular, are cleaned using a Wishab sponge.

In this instance the item is a crayon photo, and must be dusted even more gently so as not to remove the charcoal retouches.

The photograph is remounted in the original frame with the addition of a preservation-quality passe-partout to ensure that the photograph remains detached from the glass. In this instance the photograph is mounted on two different cardboard backing pieces, which are retained, and there is a good deal of silver mirroring in evidence. Localized cleaning is carried out to remove grease and dust, and reduce the effect of the mirroring to some extent.

Here, the restorer is working on an image with original colouring applied by hand, which now presents widespread areas of dotted fading. The visual unity of the image is restored by retouching with watercolours.

In this case the original passe-partout has been retained, and a shaped clearance space created between the image and the passe-partout with barrier paper.

Restoration by Lorenza Fenzi

Video made by Lorisa Andreoli

Cutting and editing by Lorisa Andreoli and Alessandra Angarano

Script: Lorenza Fenzi

Voice: Alessandra Angarano

Music: Farid Zehar and Bruno Chauveaux

MINT SERVICES

EDM Ingestion tool

http://www.youtube.com/watch?feature=player_embedded&v=K4KhrUSDCx8

DPLA-Beta Sprint Demo: MINT platform and how it can be used in DPLA

http://www.youtube.com/watch?feature=player_embedded&v=jR3HMAanmzM

Mapping Workflow 1/5: Introduction

http://www.youtube.com/watch?feature=player_embedded&v=z73GFDzjhi8

Mapping Workflow 2/5: Metadata Import

http://www.youtube.com/watch?feature=player_embedded&v=WCc86WqDUK4

Mapping Workflow 3/5: Define Root Item

http://www.youtube.com/watch?feature=player_embedded&v=w1WUK6yRof4

Mapping Workflow 4/5: New Mapping

http://www.youtube.com/watch?feature=player_embedded&v=ia5oJvOEzgs

Mapping Workflow 5/5: Publish

http://www.youtube.com/watch?feature=player_embedded&v=dX2zsv2JroY

ANNEX 2: PDF FILES REPRESENTING ALL THE LEARNING OBJECT WEB PAGES, THE ASSOCIATED READINGS AND TRAINING MATERIALS AND THE QUESTIONNAIRES (EVALUATION AND ASSESSMENT) CAPTURED BY "PRINT PAGES TO PDF" APPLICATION

Persistent identifiers: What if? (LO-01.pdf)

Digitisation life cycle (LO-02.pdf)

MINT services (LO-03.pdf)

Why and how to contribute to Europeana (LO-04.pdf)

Persistent Identifiers: Commercial and heritage views (LO-05.pdf)

Public-Private Partnership with Europeana (LO-06.pdf)

Terminology (LO-07.pdf)

Linking Cultural Heritage Information (LO-08.pdf)

The Linked Heritage Training Programme Overview

Videos Playlist on YouTube

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

ABOUT THE LEARNING OBJECT

Readings and training materials

Summary

This Learning object deals with Persistent identifiers (PIDs).
It is divided into 17 sub-units each of which consists of two parts: a dialogue between two owlets (on the left of the screen) introducing PIDs, and a video (on the right of the screen) representing the concepts.
The idea was to use the famous painting by Botticelli, *The Birth of Venus*, as a metaphor. Each PID functional requirement is represented by a visual metaphor associated with a musical metaphor: uniqueness, persistency, resolvability, reliability, authoritativeness, flexibility, interoperability and cost effectiveness.

Languages

- German
- Greek
- English
- Polish
- Spanish

Forward »

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

1/19 – INTRODUCTION

"There was a great exhibition by Sandro Botticelli in Florence only a few years ago. It was the first time so many of his works were collected in one place and I remember reading very positive reviews about it. But … I can't find the information about this event any more … I type the correct URL and what I get is *'404: not found'*. It is so annoying considering the event is so recent. I wonder what we'll find in ten years' time."

"You wouldn't have this problem if **Persistent identifiers** had been used to refer to that exhibition!"

"**Persistent?** What are you talking about? Everything is so volatile nowadays … anyway, what do you mean by 'Persistent Identifiers'?"

« Back Forward »

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

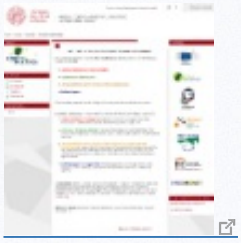
2/19 – WHAT IF: GLOBAL UNIQUENESS

"Are PIDs really necessary?"

"What if we couldn't rely on **globally unique** identifiers?"

Download HD video

Readings and training materials



« Back Forward »

Linked Heritage LEARNINGObjects

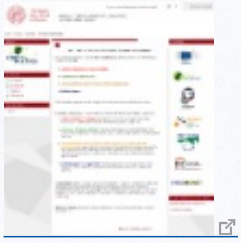
PERSISTENT IDENTIFIERS: WHAT IF?

3/19 – WHAT IF: PERSISTENCE

"What if we couldn't rely on **permanently unique** identifiers?"

Download HD video

Readings and training materials



« Back Forward »

Linked Heritage LEARNINGObjects

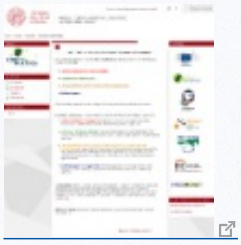
PERSISTENT IDENTIFIERS: WHAT IF?

4/19 – WHAT IF: RESOLVABILITY

"What if identifiers weren't linked to the current location of the resources?"

Download HD video

Readings and training materials



« Back Forward »

Linked Heritage LEARNINGObjects

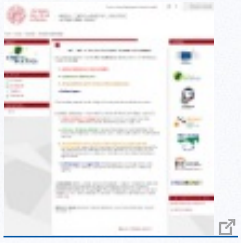
PERSISTENT IDENTIFIERS: WHAT IF?

5/19 – WHAT IF: RELIABILITY

"What if we couldn't be supported by a **reliable** PID system?"

Download HD video

Readings and training materials



« Back Forward »

Linked Heritage LEARNINGObjects

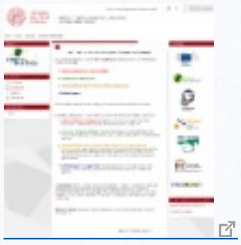
PERSISTENT IDENTIFIERS: WHAT IF?

6/19 – WHAT IF: AUTHORITATIVENESS

"What if we couldn't rely on an **authoritative** PID system?"

Download HD video

Readings and training materials



« Back Forward »

Linked Heritage LEARNINGObjects

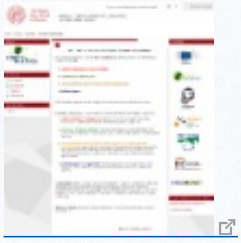
PERSISTENT IDENTIFIERS: WHAT IF?

7/19 – WHAT IF: FLEXIBILITY

"What if a PID system did not represent collections in sufficient **detail**?"

Download HD video

Readings and training materials



« Back Forward »

Linked Heritage LEARNINGObjects

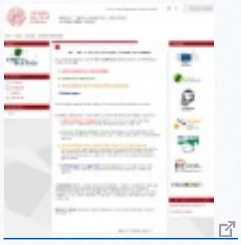
PERSISTENT IDENTIFIERS: WHAT IF?

8/19 – WHAT IF: INTEROPERABILITY

"What if we couldn't rely on a **interoperable** PID system?"

Download HD video

Readings and training materials



« Back Forward »

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

9/19 – WHAT IF: COST EFFECTIVENESS

"What if we couldn't rely on an **affordable** PID system?"

"What a mess! What happened?"

"Don't worry, you can use PIDs!
PIDs can refer to all the information associated with a real object, including its location, or to any of its potential surrogates, e.g. digital images, a museum collection where it belongs, research documents referring to it and other services. PIDs may be applied to real objects as well as to more abstract concepts like services, transformation issues, aggregation or disaggregation of objects and organizations."

« Back Forward »

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

10/19 – COST EFFECTIVE

"Due to limited funding allocated to the cultural heritage sector, cultural organisations should use PID systems that are either free or have very low costs"

COST EFFECTIVE

Resources, particularly financial resources, are scarce in the cultural heritage sector.
In addition, organisations have a general mission to provide access to their items free of charge for non-commercial use.

Therefore:

- **Cultural organisations should use PID systems that are either free of charge, or have a very low cost in relationship to their available resources.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

« Back Forward »

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

11/19 – INTEROPERABLE

"To enable the greatest number of users to share and use cultural contents, **interoperability** among different PID systems must be achieved mainly through the adoption of **open standards**."

INTEROPERABLE

This is vital to ensuring that cultural content can be shared and used by as a large a set of users as possible.
Many PID solutions were designed for specific domains.

Therefore:

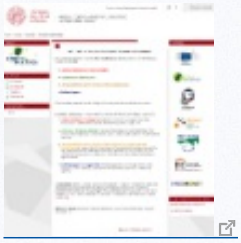
- **Organisations should use intellectually openstandards for the implementation of PIDs.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

« Back Forward »

Download HD video

Readings and training materials



Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

12/19 – FLEXIBLE

"An identifier system should adjust to different requirements of different types of collections, for example it should be able to deal with varying levels of details, from individual objects to aggregations: the granularity of collections should be represented. This is called **flexibility**."

FLEXIBLE

A PID system will work more effectively if it can handle the requirements of different types of collections.
Parts of collections may be curated at different levels of 'granularity', from parts of objects, to individual objects, to collections objects. The latter has an unbounded number of individual elements.

Therefore:

- **Organisations should use PIDs systems that are flexible enough to represent the granularity of their collections.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

« Back Forward »

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

13/19 – AUTHORITATIVE

"Organisations should establish the **authority** and credibility of a PIDs system's provider before adopting that system."

AUTHORITATIVE

Some PID systems are dependent on responsible organisations who: manage the system, assign identifier; and resolve the identifiers to resources.
Some services are provided by public institutions like national libraries and archives.
For a system to be effectively supported the responsible organisation must be able to demonstrate its commitment.

Therefore:

- **Organisations should evaluate and be assured of the authority and credibility of a PIDs system's provider before adopting that system.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

« Back Forward »

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

14/19 – RELIABLE

"The **reliability** of PID systems should be certified by institutions by means of redundant technology and register updates (preferably automatic)."

« Back Forward »

RELIABLE

For a PIDs system to function reliably these issues have to be assessed:

1. It should always be active (e.g. backed up, with redundant technology)
2. The register of PIDs should be updated (preferably automatically)

Therefore:

- **Organisations should evaluate and be assured of the technical reliability of a PID system (including their own) before adopting it.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

15/19 – RESOLVABLE

"... you see ... the location has changed, how lucky I am I can still trace it!"

"Yes you are! As a matter of fact **resolution** services guarantee that PIDs are linked to their current resource URLs even when the latter change.
So, don't mix up resolution with retrieval, which is the ability of systems to access and download digital resources by clicking on URLs."

« Back Forward »

RESOLVABLE

Choice to use PIDs does not imply that an external human user will be able to access anything that they can use effectively.

Therefore:

- **Organisations should be clear, and make public, information about which, if any, their PIDs resolve to an available resource.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

16/19 – PERSISTENT

"You see, **persistence** refers to a number of aspects which are not only technical, but also related to the policies and commitments of institutions. They should be committed to creating and maintaining their PIDs in order to give exhaustive information about their collections, no matter what."

« Back Forward »

PERSISTENT

Persistence refers to the lifetime of an identifier, during which it should not possible to reassign it to another resource or to delete it.
If an organisation can guarantee that a PID will be managed so that it will survive changes to ownership and PID system, then an external user can be confident of its persistency.

Therefore:

- **Organisations should commit themselves to the persistence of their PIDs and make clear to others what they mean by 'persistent' and how this will be implemented.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

Download HD video

Readings and training materials

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

17/19 – UNIQUENESS ENVIRONMENT

"A PID is a label associated with something in a particular environment. On the Internet it should be **globally unique**, but it may only be so within a specified context. At any rate, it must be unique within the institution's own system."

"Oohhh, what a masterpiece!"

« Back Forward »

UNIQUENESS ENVIRONMENT

A PID is label that is associated with something in a particular environment.
On the Internet it should be globally unique, but may only be unique in combination with a limited name space. In the 'worse' case it may only be unique within an organisation's own systems.

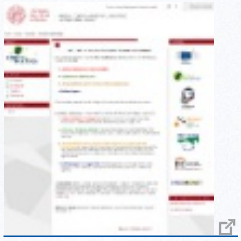
Therefore:

- **Organisations should be clear, and make public, in which environment their PIDs are unique.**

Persistent Identifiers (PIDs): recommendations for institutions (PDF) by Gordon McKenna and Roxanne Wyns.

Download HD video

Readings and training materials



Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

18/19 – CONCLUSIONS

Thank you for viewing our Learning Object *Persistent Identifiers: What if?*

We invite you now to answer some questions for your learning assessment and to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: PIDs: What if.

For further details:

Readings and training materials

« Back Forward »

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: WHAT IF?

19/19 – CREDITS








Project by:
Cristiana Bettella
Elena Bianchi
Lisetta Dainese
Massimo Ferrante
Michele Visentin
Antonella Zane

CAB, University of Padua (Italy)

Texts based on the Athena project booklet *Persistent Identifiers (PIDs): recommendations for institutions* (PDF) by Gordon McKenna and Roxanne Wyns.

Graphic and Web design by Gianluca Drago
CAB, University of Padua (Italy)

Technical support and Flash development by Angelo Calò, Cecilia Dal Bon
CMELA, University of Padua (Italy)

Audio by Lodovico Bollacasa, Padua (Italy)

Music from:
Wolfgang Amadeus Mozart, *Piano Sonata No.16 in C major, K.545, "Sonata facile"*
Sergej Sergeevič Prokof'ev, *Peter and the Wolf*
Frederic Chopin, *Valtz in A flat Major op.69 n.1*
Lodovico Bollacasa, *Petite fille*

Acknowledgments:

We would like to thank Andrea Belli, Valeria Donata Bettella and Lodovico Bollacasa (Milan – Padua, Italy) for the interoperability images, and the pianist Ugo Armano (Padua, Italy) for performing Mozart's *Sonata facile*.
A special thank to Raffaella Traniello (multimedia projects consultant) for her fundamental contribution to the design of the Linked Heritage Learning Objects.

Funded by the Competitiveness and Innovation Framework Programme, 2011.

« Back

Linked Heritage LEARNINGObjects

DIGITISATION LIFE CYCLE

ABOUT THE LEARNING OBJECT

Readings and training materials

Summary

This Learning Object presents the digitisation workflow both in theory and in practice. It consists of two parts: the first one gives an overview of the digitisation entire workflow, the second one focuses on a case study of the University of Padua: the "Botanists portrait collection".
The case study is illustrated also by means of two videos showing the following activities: preservation of items and digitisation.

Languages

- English
- German
- Greek
- Italian
- Spanish

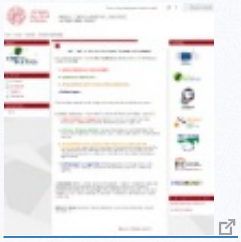
Forward »

Linked Heritage LEARNINGObjects

DIGITISATION LIFE CYCLE

1/5 – INTRODUCTION

Readings and training materials



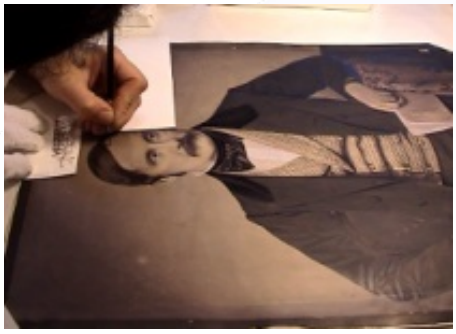
Workflow

The purpose of this resource is to delineate the range of procedures for the digitisation of two-dimensional documents, in conformity with best practices and with national and international standards for quality reproduction of documents.

In essence, digitisation projects should be consistent with one or more of the following general aims:

- deriving maximum possible advantage from documents preserved in archives, museums and libraries
- widening public access to documents and their content for scientific and cultural purposes
- facilitating interdisciplinary studies and promoting collaboration between different institutions
- promoting awareness of local or unique collections, through large-scale dissemination of their content
- creating virtual collections through the integration of different formats or materials distributed in different locations
- limiting direct consultation of original documents in particularly critical conditions
- facilitating access to typically inaccessible material
- ensuring that documents will be made available to future generations of students, scholars and members of the general public

A case study at the University of Padua



The Botanists portrait collection includes 2,380 portraits of Italian and foreign botanists from the second half of the 17th century to the first half of the 20th century. The collection is made up of photographs (salted paper prints, albumen prints, aristotypes, platinum prints, gelatin silver prints), glass negatives, engravings, watercolours, drawings, paintings and photomechanical prints. The collection is preserved in the Padua Botanical Garden Library.

In 2005, the Botanical Garden Library and the Ancient and special collections section of the Padua University Library System initiated a project designed to derive maximum benefit from the collection. The aims of the project are to ensure long-term preservation and on-line access of the catalogued and digitised collection. The project envisages six steps:

inventory and selection, legal aspects, preservation, cataloguing, digitisation, on-line access.

EXPLORE

Information of the Botanists portrait collection

Examples from the Botanists portrait collection: Nadar's photographs:

- Digitised images in Phaidra:
  - Giovanni Antonio Maria Zanardini
  - Pierre Edmond Boissier
- Bibliographic records from On-line Public Access Catalogue (OPAC) of the Padua University Library System:
  - Giovanni Antonio Maria Zanardini
  - Pierre Edmond Boissier
- Nadar: authority record

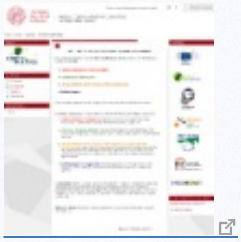
« Back Forward »

Linked Heritage LEARNINGObjects

DIGITISATION LIFE CYCLE

2/5 – WORKFLOW

Readings and training materials



Page Index

Selection

The selection of documents will normally take into consideration the growth of the collection, added value, protection, technical feasibility, and ability to sustain the long term costs of digitisation. In reality, the best selection will be based on a combination of criteria.

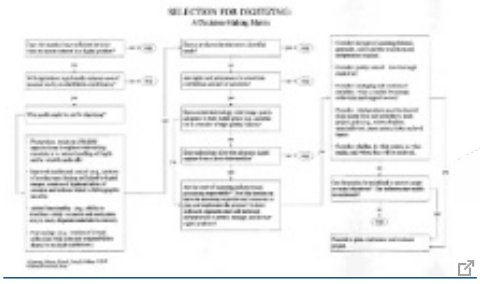
Selection criteria will generally take account of:

- historic and cultural value
- uniqueness and rarity
- high demand
- documents free of legal constraints, or with permission to digitise already secured
- limited access due to state of preservation, value and location
- value added by providing access on line; creation of virtual collections
- increased level of interest generated in little-known or unknown material

In certain cases it may be worth conducting a survey of the material so as to determine the quantity, type and format of the documents, and their state of preservation. This information may come in useful during subsequent preservation, cataloguing and digitising operations.

EXPLORE

Selection for digitising: a decision-making matrix



Top

Legal aspects

When digitising documents, serious attention must be given to issues concerning copyright, in respect both of original material and of digital resources.

Points to examine are: characteristics of the work to be processed, rights ownership (who owns the rights – is the work protected – what type of protection?), the actions to be performed on the work (what are they – what rights are involved – has authorization been obtained?), likely critical areas and possible solutions.

Works that must be excluded are those subject to copyright and those digitised in other collections and accessible to the public on the web, in this instance so as to avoid duplication and minimize costs.

EXPLORE

Arrow
IPR Helpdesk
Step-by-step IPR Guide

Top

Preservation of items

Digitisation is no substitute for commitment to care and preservation of original documents. It is important to assess the state of preservation of original documents before proceeding with digitisation, and to ensure that any treatment of original specimens is carried out only after they have been inspected by experts.

EXPLORE

IFLA Principles for the care and handling of library material
The Library of Congress, Preservation – Collections care resources
Northeast Document Conservation Center (NDCC) – Resources: Preservation leaflets

Top

Digitisation

To guarantee the safety of originals and ensure good digitising quality, particular care must be taken over the choice of acquisition methods and equipment (capture system, lighting, software).

The nature and the dimensions of the originals will determine the selection of the capture system and the lighting system.

The hardware and software requirements of the capture system are determined by the image quality expectation, as are the timescales for acquisition and processing of the images, and the amount of storage space occupied in allocated memory locations.

As a general rule, the key to quality does not lie in scanning at the maximum resolution obtainable, but performing the scan at a level commensurate with the information contained in the original.

In general, the aim of digitisation is to produce master files suitable for long-term preservation. Files for viewing on the web are derived from the master files.

In-house or outsourced digitisation

The decision as to whether documents should be digitised by the institution (in-house), or entrusted to an outside service provider (outsourcing) will depend on the advantages and drawbacks of the two methods.

	In-house	Outsourcing
Advantages	• keeping direct control over the entire process • learning by doing • improving standards as work proceeds, rather than setting targets beforehand • ensuring safety, proper handling and accessibility of materials	• institution pays for the end product, generally on the basis of an agreed price per image • costs are kept down, and risks limited • service provider can handle large quantities of material • rovider shoulders the costs of specialization, training and technological obsolescence • wide range of options and services available
Drawbacks	• rather than paying for the product, institution shoulders the costs, including training, technological obsolescence and down time • outlay on purchase and maintenance of equipment • need for skilled human resources • cost per image not defined	• by eliminating a step of the process, institution does not develop thorough knowledge of digitisation • problems with safety, transportation and handling of original specimens
Recommendations	In-house is best if: • the collection cannot be moved outside of the institution • the digitising process is a very simple one • reliance can be placed on specialist human resources and equipment already on site	Outsourcing is best if: • original specimens cannot be digitised in-house for whatever reason • schedule involves processing large quantities of material in the short term • there are constraints in terms of space, infrastructure and human resources

If the decision is made to **entrust the service to a company**, the institution must:

- determine the digitisation parameters
- draw up a detailed invitation to bid
- evaluate the products and services offered
- define the contractual responsibilities of the institution and of the company
- carry out a final quality control on the product

The **cost of digitisation** depends on a number of variables, namely the size, type and nature of the document for digitisation and the envisaged use of the digital object, hence an assessment of the costs can be requested from the digitisation service provider, or alternatively, based on previous digitisation projects. It may help to consult existing literature

on the topic.

Selection of equipment

General indications on the capture system:

- **Flatbed scanners** are used for single sheet documents, or bound documents that can be opened out flat without difficulty, of dimensions up to paper size A3 (420 x 297 mm). These documents include: printed matter (e.g. flyers, posters, brochures), manuscripts (e.g. letters), maps in good condition, sheet music, prints (e.g. engravings, etchings, lithographs), pen-and-ink drawings with no added water colour or tempera (e.g. cartoons), photographic material (e.g. black-and-white and colour gelatine prints, albumen prints).
- **Scanners for films and transparencies** are used for films, negatives and transparencies.
- **Planetary scanners or Digital cameras** are used for bound documents, documents of a special nature, and documents larger than size A3.
  These documents include: bound volumes (e.g. books, albums, sheet music, atlases), fragile documents, oil paintings, most works of art on paper (e.g. watercolours, drawings), graphic material and works of art created with flaky and friable substances (e.g. crayons, charcoals, soft pencil), watercolours applied thickly, with tempera or varnishes, large or fragile maps, manuscripts (e.g. bound diaries, folded documents), parchments, photographic material (e.g. large size prints; historic photographic processes such as daguerreotypes or ambrotypes), three-dimensional material (e.g. fabrics, sculptures, objects).

In the case of antique and fine art originals, the lighting system must be fitted with lamps emitting cold light and ultra-low levels of IR and UV radiation.

Digital acquisition

Bearing in mind the resources available, the decision on image quality should be based on the needs of users, on the method of delivery and use of images, and on the nature of the materials being digitised (size, format, type of material, colour, etc.).
There are various reasons for creating a high quality master: preservation, access and cost, and the need to ensure that the digitisation process will not need to be repeated in future. The master can be used to prepare files in smaller sizes or alternative formats for the different uses envisaged. Standard formats should always be used.

Indications on the master file:

- this is the file in which the single digital object is created and preserved, and from which derivatives can be generated (JPEG, PDF etc.); enables high quality printing
- the master file represents the informative content of the original, as closely as possible
- the original must be captured in its entirety. A border must be left around the document, so that the outline of the image can be identified
- if the original is mounted on a backing that carries information, the digitisation should also include the backing
- the master file is archived exactly as reproduced by the acquisition tool
- the file should be in a standard format, such as TIFF
- the title of the file should incorporate a colour profile
- if the original is digitised and accompanied by colour scale, grey scale and gauge, these shall be located outside the borders of the reproduced image and within the overall perimeter of the surround

Indications on derivative files:

- these are used in place of the master for the purposes of LAN or WAN access, and accordingly, the dimensions will depend on the envisaged uses
- derivative files should be of suitable proportions for fast download, without requiring a high speed connection, of acceptable quality for general research purpose, and presented in a compressed format for speed of access
- the usual formats are JPEG or PDF

File naming

Before commencing any acquisition procedures, file-naming criteria must be established. In general terms, the name of each file will consist in a string of characters that must contain the information needed to identify the element of the collection to which the image belongs, uniquely and unambiguously. Filenames will be completed with the appropriate extension, such as ".tif", ".jpg", etc.

Storage of data

The collection of the images, consisting in directories and files, will be memorized on optical or magnetic storage media such as CDs, DVDs, and external hard disks.
It is important that data should be saved to at least two such elements of storage media, preserved at two distinct locations, and that the data should be checked and refreshed periodically. The life of the storage media is in any event influenced by a variety of factors (ISO 18923:2000 and 18925:2002 standards indicate the parameters for proper preservation of storage media).

Quality control

Quality control should be documented and conducted throughout the entire digitising process on all material captured, and in particular on master files.

Planning of the quality control system should include:

- appropriate preparation of the environment (hardware configuration, viewing software, viewing conditions, etc.)
- *a priori* definition of "acceptable" and "not acceptable" characteristics
- verification procedure (entire collection or sample, all files or master files only, visual quality on screen, in print, etc.)

EXPLORE

California Digital Library (CDL) - Guidelines for Digital Images
Cornell University Library - Moving theory into practice: digital imaging tutorial
Digital Library Federation (DFL) - Draft benchmark for digital reproductions of printed books and serial publications
Digital Library Federation (DLF)- Guides to quality in visual resource imaging
Federal Agencies Digitisation Initiative (FADGI), Still Image Working Group - Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files
S. Hubbard, D. Lenert (edited by) - Introduction to Imaging
National Archives and Records Administration (NARA) - Technical guidelines for digitising archival materials for electronic access: creation of production master files – raster images
Northeast Document Conservation Center (NDCC) - Handbook for digital projects

Top

Metadata

Metadata is structured information relating to any type of resource, used to identify, describe, manage or allow access to the resource in question.

There is no metadata standard that meets all the needs of all types of collections and repositories.
Generally considered, metadata models include the following information:

- Descriptive metadata: data describing the content of a resource and allowing its retrieval
- Administrative metadata: data containing information on the management and

administration of a resource (e.g. rights management, preservation metadata, technical metadata)

- Structural metadata: data describing the relations between digital objects (e.g. page order in a digitised book)

From *Good practices handbook* (edited by the Minerva Working Group 6)

> ***"Appropriate Meta-data Standards***
>
> ***Issue Definition***
>
> *Certain important standards already exist for meta-data. In the bibliographic domain (and increasingly in non-library cultural domains), the Dublin Core standard is of great importance.*
>
> ***Pragmatic Suggestions***
>
> - *Review existing meta-data models and standards before creating your own.*
> - *Creating a totally new meta-data model for cultural collections should be avoided.*
> - *The meta-data work carried out by similar projects in the past is likely to be relevant to your project – meta data models travel well between projects in the cultural area.*
> - *Unless your project has good reason not to do so, the Dublin Core fields should be included in the meta-data model. While museums may find the CIMI model better fits their holdings, a common core set of attributes should be aimed for, which will enable cross-collection searching.*
> - *If a proprietary meta-data model is to be used, a mapping from this model to the Dublin Core should also be developed.*
> - *While a naming scheme or national naming convention may be very useful, a full meta-data model is better, both in terms of the amount of data that can be stored about an item, and also to enable more powerful searching and interoperation with other projects and other countries."*

EXPLORE

ATHENAWP3 (edited by) - Digitisation Standard Landscape
M. Baca (edited by) - Introduction to metadata
Dublin Core Metadata Initiative Wiki - User guide
JISC Digital Media - An introduction to metadata
JISC Digital Media - Medatada standards and interoperability
National Information Standards Organization (NISO) - Understanding metadata

Top

Digital preservation

In any digitisation project, it is essential to maintain digital resources created over time in such a way that burdensome repeats of digitising operations will be avoided. Accordingly, the institution must adopt procedures to ensure that digital objects will remain usable and accessible, irrespective of technological changes in the future.

The usability and accessibility of digital objects over time is guaranteed by the file format (standard for formats, file sizes, web transmission rate, methods of viewing images…), and by the archiving media and digital repository (digital objects with associated metadata will be archived and managed in a digital repository). It is fundamentally important to use open standards, thereby facilitating interoperability with other systems, and allowing access to metadata through other service providers (e.g. Europeana).

EXPLORE

Cornell University Library - Digital preservation management resource
B. F. Lavoie - The Open Archival Information System Reference Model: Introductory Guide

Top

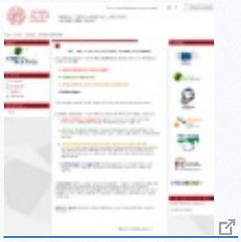
« Back Forward »

Linked Heritage LEARNINGObjects

DIGITISATION LIFE CYCLE

3/5 – A CASE STUDY AT THE UNIVERSITY OF PADUA

Readings and training materials



Page Index

Selection

Making an inventory of the collection was the first essential step in assessing the quantity, type and size of the items, and their state of preservation.

A survey of the material produced the following data: inventory, location, name of botanist (birth-death), specific designation of material (e.g. postcard, photograph, painting…), method of creation, dimensions, type of material and dimensions of mount, state of preservation, information on duplicates, if any, and notes.

The selection process involved the entire collection, for the following reasons: historical value and uniqueness, preservation, increased access to content, improvement of services and absence of legal constraints.

Top

Legal aspects

The documents in the collection are subject to Italian Law n° 633 of 22 April 1941 “Protection of copyright and other rights relating to the exercise thereof” and subsequent amendments.

The legal aspects taken into consideration are:

- **Portrait of a person**: in the case of the botanists portrait collection, a search for the heirs of the persons portrayed is impractical. For the publication of images, reference is made to the exception of article 97, Law n° 633/1941 and subsequent amendments.
- **Copyright**: photographs represent 70% of the collection, and in many cases are reproductions of other photographs or other works of art. It is important to highlight the distinction between photography as an *intellectual property* and photography as a *simple reproduction of an existing reality*. The first is an exclusive right and ownership is protected for 70 years following the death of the author; the second is a connected right and ownership is protected for 20 years from the date on which the photograph was produced.
  In any event, all works come into the public domain once 70 years have elapsed following the death of the author, and this is the status of all images in the collection.
- **Acquisition**: there is no formal deed by which the collection was donated; it was Pier Andrea Saccardo (prefect of the Botanical Garden from 1879 to 1915) who had the idea of collecting portraits of botanists. Images have been donated to the Library by private

individuals and by institutions.

EXPLORE

Italian Law n° 633 of 22 April 1941 “Protection of copyright and other rights relating to the exercise thereof”

Top

Preservation of items (video)

The preservation of the portrait collection involved the following operations: dusting and housing of specimens with materials to ISO standard, restoration of negatives and framed photographic prints, and restoration of prints.

Video transcription (English translation):

Housing and photographic restoration of the Botanists portrait collection

Housing

The item is removed carefully from the original envelope. First, any loose dust is blown away using a rubber bulb syringe, then both sides are cleaned with a

To turn captions on, start video and click the small icon in video player

Download video in HD

Download transcription (English translation)

Dusting and housing

Items were treated adopting the following procedure:

- removal of items from boxes and from original envelopes
- removal of dust particles from the image with a rubber bulb syringe, and cleaning with a soft bristle brush
- housing of documents in their envelopes or in folders suitable for preservation. Envelopes were placed in boxes to ISO standard
- return to storage: boxes in a cabinet, and folders in a chest of drawers

The quality of the material is in accordance with ISO 10214 (Photography – Processed Photographic Materials – Filing Enclosures for Storage) and ISO 14523 (Photography - Processed Photographic materials - Photographic activity test for enclosure materials).

Restoration

Glass plate negatives and framed photographic prints were entrusted to a professional photograph restorer; prints were entrusted to a paper restorer.

EXPLORE

Photographic documentation on preservation of the portrait collection
Il Progetto di digitalizzazione "Iconoteca dei botanici", p. 32–38

Film about restoration of prints in the portrait collection

Information resources on treatment of photography:

- George Eastman House - Photographic Processes and Terms Glossary
- The Library of Congress - Preservation - Care, Handling, and Storage of Photographs: information leaflet
- State Library & Archives of Florida - Daguerreotype to Digital: A Brief History of the Photographic Process

Top

Digitisation (video)

In accordance with digitising guidelines and best practices, parameter selection covered the size, type, information content and envisaged use of the digital object.
The aim of the digitisation process was to produce master files for preservation; files derived from these master files would then be made available for consultation on local and wide area networks.

In-house or outsourced digitisation

The operation of digitising the Botanists portrait collection was outsourced for the following reasons: the collection is relatively small (2,380 portraits); the Library possesses neither the appropriate equipment, nor the necessary specialist human resources; the work had to be completed by 2008.

The digitisation process was carried out on the Library premises, to avoid the disruption of moving the items and the associated problems of insurance.

The invitation for tenders to provide the digitisation service was conducted in line with Padua University regulations, based on the *Specification for digitisation activity* of OTEBAC (Osservatorio Tecnologico per i Beni e le Attività Culturali – Technological Observatory for Cultural Heritage and Activities), and on the *Request for proposal* of the NDCC (Northeast Document Conservation Center) and the RLG (Research Libraries Group).

Selection of equipment

The Botanists portrait collection is heterogeneous in terms of the type, dimensions and fragility of the various specimens, and accordingly, it was decided to use a digital photo capture system requiring no contact with the original, comprising:

- Hasselblad CF39 MS digital back with 39 million pixel sensor and control station
- Hasselblad 500 ELX professional cameras with Planar 80 mm and 120 mm lenses, also PK 35 mm wide angle and Nikon AF 60 mm micro lens
- HMI lighting system providing cold light and flat beam
- Backlit slide viewer for transparencies

Digital acquisition

A. General indications for the master file:

- the original is photographed/scanned in its entirety, with no cuts. A border of approximately 4 mm (or in any event approximately 8 pixels), preferably white, is left around the document.
- if the original is applied to a mount or support carrying information (e.g. photograph in visiting card format), the mount must also be digitised
- geometric accuracy must be observed
- the master file is archived exactly as reproduced by the acquisition tool
- the digitised image of the original must be accompanied by colour scale, grey scale and gauge
- the scales and gauge must not overlap the outline of the reproduced images
- the pixel proportion must be square
- the digitised image must be devoid of reflections, especially in the case of photographs, engravings and framed pictures

B. Requirements for size, resolution, format and compression of files:

Resolution
The effective optical resolution of the captures is tied to the physical size of the originals (the set), as indicated in the following table:

Set	Measurements	Min PPI
1	cm 5x3,75	3664 PPI
2	cm 10x7,50	1832 PPI
3	cm 20x15,01	916 PPI
4	cm 30x22,51	611 PPI
5	cm 40x30,02	458 PPI
6	cm 50x37,52	366 PPI
7	cm 60x45,02	305 PPI

A dedicated post-processing program was defined, with functions that included automatic cropping and automatic positioning on a white background (RGB 255, 255, 255), with a surround corresponding to 102-3% of the dimensions of the object. The automatic process

produced a quantity of digital objects which, according to the specific nature of the originals, required a variety of manual treatments.

Format

- Master file for archiving: TIFF uncompressed, colour depth 16 bits per channel (48 bit RGB colour; 16 bit greyscale for negatives), IBM PC byte order, interleaved pixel order, ProPhoto RGB colour profile, with optical resolution depending on the dimensions (see table)
- File derived from master for consultation in LAN or at high speed: JPEG compressed, maximum quality (Adobe Photoshop scale 12) colour depth 8 bits per channel (24 bits RGB), cut out without gauge/Colour, sRGB colour profile to IEC-61966-2.1, resolution 300 ppi
- File derived from master for consultation on Web: JPEG compressed, high quality (Adobe Photoshop scale 8) with colour depth 8 bits per channel (24 bits RGB), cut out without gauge/Colour, sRGB colour profile to IEC-61966-2.1, resolution 96 ppi
- Thumbnails: JPEG compressed medium quality (Adobe Photoshop scale 5) with colour depth 8 bits per channel (24 bits RGB), cut out without gauge/Colour, sRGB colour profile to IEC-61966-2.1, resolution 96 ppi, width fixed at 150 pixels, height proportional

Video transcription (English translation):

University of Padova
University Library System

Antique and special collections section
Botanical Garden Library

Digitisation of the Botanists portrait collection

January 2009

To turn captions on, start video and click the small icon in video player

Download video in HD

Download transcription (English translation)

File naming

File identification was accomplished using the native algorithm from the software owned by the company, already employed in projects of the BDI (Biblioteca Digitale Italiana – Italian Digital Library). The algorithm uses 19 characters plus 4 for the file extension; the encoded characters contain information on the medium, the univocal identifier and the file version.

Storage of data

All files are archived on a server of the CAB (Centro di Ateneo per le Biblioteche – University Libraries Centre) and on two external hard disks, one copy on each. The two hard disks are kept at the Botanical Garden Library and at the Centre (CAB).

Quality control

Using an on-line link to the consulting system of the contracted company, it was possible to check the files during the course of the digitising process and report any defects. Defective files were re-digitised.
Master files underwent sample inspections to check for:

- correct frame alignment and exposure, absence of possible deformations and/or optical aberrations
- containment of chromatic tolerance

- colour depth and profile
- digital size and format
- possible existence of elements having an adverse effect on fidelity of reproduction (e.g. dirt, light reflections, etc.)

In the case of files derived from the master files, the inspection was extended to all available items in order to verify the general quality, and the presence of defects, if any, within and immediately adjacent to the images.

Top

Metadata

Cataloguing of items

The collection is catalogued in accordance with ISBD(NBM) standard. Authors are identified in accordance with RICA (Regole italiane di catalogazione per autori – *Italian cataloguing rules for authors*.
Cataloguing follows the UNIMARC bibliographic format and was done in Aleph, the management automation system used by the SBA (Sistema Bibliotecario di Ateneo – University Library System).

UNIMARC FIELDS

Show UNIMARC fields

Metadata of digital objects

The metadata schema selected was MAG1 (Metadati Gestionali e Amministrativi – *Management and Administrative Metadata)* – version 2.0.1 - by ICCU (Istituto Centrale per il Catalogo Unico – Central Institute for the Single Catalogue of Italian libraries and bibliographic information).
The sections used are: GEN, BIB, STRU and IMG.

To generate the BIB section, descriptive items of metadata in UNIMARC format were exported by Aleph to a single ISO 2709 file and supplied to the company for importation into its operational database.

Top

Digital preservation

The Botanists portrait collection is archived in Phaidra (Permanent Hosting, Archiving and Indexing of Digital Resources and Assets), the management system for digital objects with long-term archiving functions used by the SBA (Sistema Bibliotecario di Ateneo – University Library System). Phaidra is Europeana compliant.

Publication on-line using Phaidra involves the following workflow: mapping of MAG / Phaidra metadata, script for bulk upload of the collection, creation of collections for obverse/reverse digitised images, trial upload on Phaidra test, migration of data from Phaidra test to Phaidra production and mapping between Aleph system numbers and Phaidra URL so that the URL can be included automatically in bibliographic records, thereby creating a mutual link between bibliographic records and digital objects in the collection.

Top

Bibliography

Andreoli, Lorisa, *L'Iconoteca dei botanici e il fotografo Luigi Caporelli*, "AFT : semestrale dell'Archivio Fotografico Toscano", n. 47, a. XXIV (giugno 2008)

Id., Il Progetto di digitalizzazione "Iconoteca dei Botanici" Rapporto tecnico, Padova, luglio 2011

Beguinot, Augusto, *I materiali di archivio del r. Istituto ed Orto Botanico di Padova,* "Boll. dell'Ist. Bot. della R. Univ. di Sassari", vol. 1, mem. X (1922)

Minelli, Alessandro (a cura di), *L'Orto botanico di Padova 1545-1995,* Venezia, Marsilio, 1995, pp. 311

Saccardo, Pier Andrea, *La Botanica in Italia: materiali per la storia di questa scienza*, Vol. 1, Venezia, Tip. Carlo Ferrari, 1895, pp. 236

Id., *La Botanica in Italia: materiali per la storia di questa scienza,* Vol. 2, Venezia, Tip. Carlo Ferrari, 1901, pp. 172

Id., *La iconoteca dei botanici nel r. Istituto botanico di Padova,* Genova, Tip. Ciminago, 1899, pp. 35 (estr. da: Malpighia, a.13, v.13)

Id., *La iconoteca dei botanici del r. Istituto botanico di Padova. Supplemento,* Genova, A. Ciminago, 1902, pp. 22 (estr. da: Malpighia, a.15, v.15)

Settimana della cultura scientifica 2008: filmato sulla mostra *Volti e luoghi dei botanici (Faces and places of botanists)* at the Botanical Garden Library

« Back Forward »

Linked Heritage LEARNINGObjects

DIGITISATION LIFE CYCLE

4/5 – CONCLUSIONS

Thank you for viewing our Learning Object *Digitisation Lyfe Cycle*

We invite you now to answer some questions for your learning assessment and to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: Digitisation Lyfe Cycle.

For further details:

Readings and training materials

« Back Forward »

Linked Heritage
LEARNINGObjects

DIGITISATION LIFE CYCLE

5/5 – CREDITS

Project by:
Cristiana Bettella
Elena Bianchi
Lisetta Dainese
Massimo Ferrante
Michele Visentin
Antonella Zane

CAB, University of Padua (Italy)

Texts by Lorisa Andreoli (CAB, University of Padua) based on the documents: *Linee guida sulla digitalizzazione*, Phaidra Working Group, by Lorisa Andreoli and Marina Cimino, May 2011 (Phaidra » Guida alla digitalizzazione); *Il Progetto di digitalizzazione Iconoteca dei Botanici*, by Lorisa Andreoli, July 2011.

Graphic and Web design by Gianluca Drago
CAB, University of Padua (Italy)

Preservation video: filming by Lorisa Andreoli, video editing by Lorisa Andreoli and Alessandra Angarano (CAB, University of Padua), text by Lorenza Fenzi (photograph restorer, Bologna, Italy), voice by Alessandra Angarano, music by Farid Zehar and Bruno Chauveaux.

Digitisation video: filming by Lorisa Andreoli, video editing by Lorisa Andreoli (CAB, University of Padua).

Acknowledgments:

We would like to thank Gabriella Valentini (photographer, Trento) author of subtitles in the digitisation video, Lorenza Fenzi author of subtitles in the preservation video.

Funded by the Competitiveness and Innovation Framework Programme, 2011.

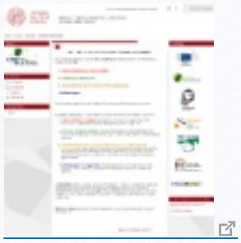
« Back

Linked Heritage LEARNINGObjects

MINT SERVICES

ABOUT THE LEARNING OBJECT

Readings and training materials



Summary

This Learning Object is for technicians (librarians, archivists, museum curators) involved in metadata mapping procedures. It could be a useful resource also for managers of cultural institutions because they can gauge the necessary skills their staff should have and decide if they can contribute to Europeana independently or with the help of an aggregator. The resource comprises 4 sections: Aggregators, the Linked Heritage project aggregator, the MINT platform, and Practical tips. The mapping workflow in MINT – the technological platform developed by the National Technical University of Athens – is described in detail in the training materials and further readings, and visualised through screencasts. The Learning Object also gives practical tips to technicians working on the mapping activities.

Languages

- English
- Spanish

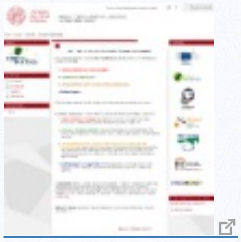
Forward »

Linked Heritage LEARNINGObjects

MINT SERVICES

1/6 – AGGREGATORS

Readings and training materials



Page Index

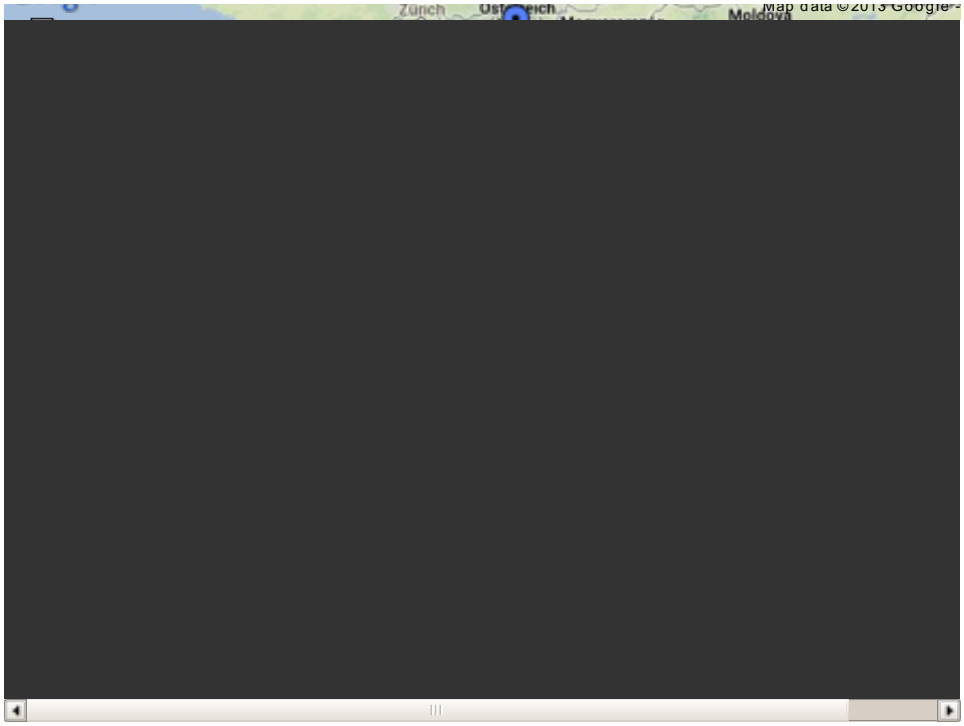
The aggregation landscape: Europeana and the European content aggregators' role

Aggregate is one of the four strategic tracks (*aggregate*, *facilitate*, *distribute* and *engage*) that underpins Europeana's future direction and business success. It is one of the means by which Europeana aims to guarantee:

- Source content that represents the diversity of the European cultural heritage
- Extend the network of aggregators
- Improve the quality of metadata

A national aggregation initiative in every EU member state is Europeana's aim by 2015.

National initiatives and aggregators – Google Maps
The map below shows each of the European aggregators that collect data from local institutions to distribute the metadata directly to Europeana.



Show National initiatives and aggregators in a bigger map.

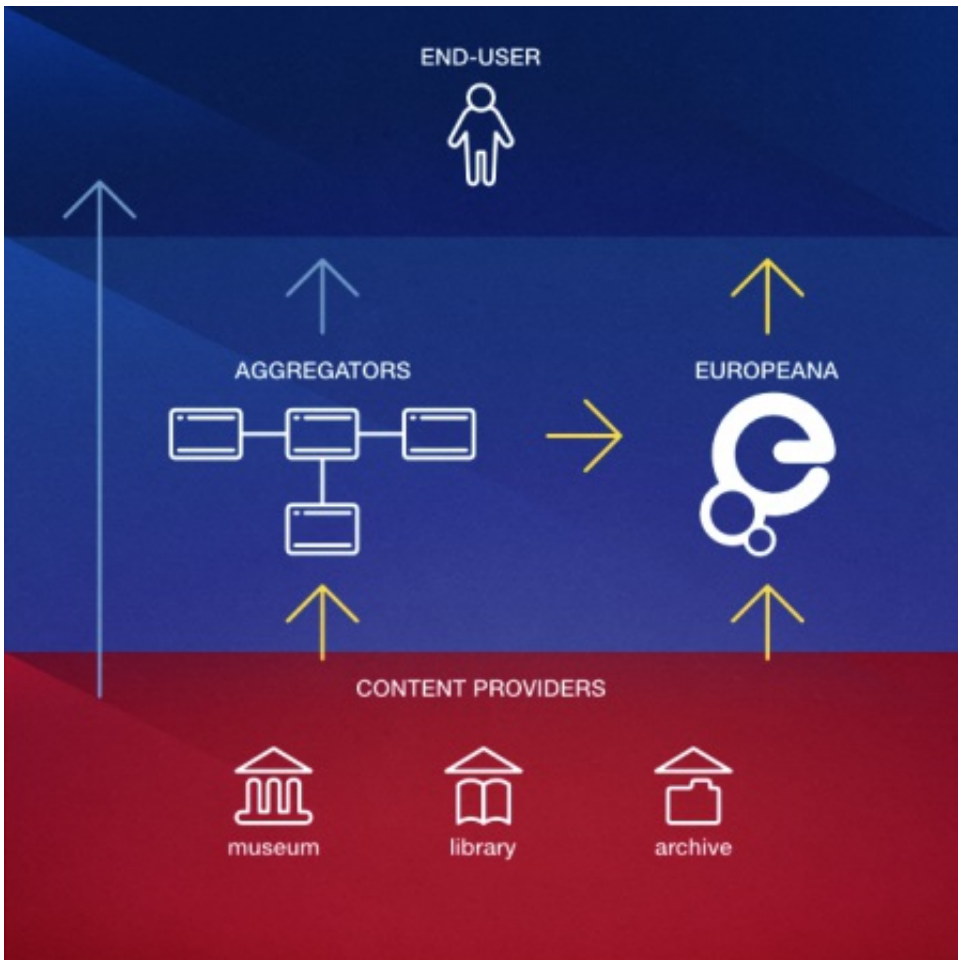
AGGREGATOR

An aggregator in the context of Europeana is an organisation that collects metadata from a group of content providers and transmits them to Europeana.

CONTENT PROVIDER

A content provider (or data provider) in the context of Europeana is any organisation that provides digital content accessible via Europeana. Europeana only ingests and indexes the institution's metadata, while the digital objects remain by the original institution.

A full list of Europeana's content providers is available at http://www.europeana.eu/portal/europeana-providers.html



Top

Aggregators: what are they?

Functions

Aggregators **gather** material from individual organisations, **standardise** formats and metadata, **channeling** them into Europeana according to Europeana guidelines and procedures.

Aggregators also **support** content providers with **administration**, **operations** and **training** (see: Linked Heritage Project Aggregator).

Roles and responsibilities

Further key roles that aggregators play in other fields are the following:

- Disseminating the vision and the objectives of Europeana to their network of institutions in order to increase support for and involvement with Europeana
- Providing valuable feedback about the issues and discussions from their field
- Promoting and implementing standards farther along the content provision chain
- Providing domain-specific expertise and skills to institutions and Europeana

Types of aggregators

- Country Aggregators
  - Regional aggregator
  - National initiatives
  - National aggregator
    CulturaItalia
- Project Aggregators
  Athena
  Linked Heritage
- Independent Organisations

Aggregators representation

DOMAIN	**Cross**	Aggregate data across domains *e.g.* Erfgoedplus.be
	Single	Data from a single domain at regional, national or international levels *e.g.* Apenet Project
	Thematic	*e.g.* Collections of Jewish culture project Judaica
GEOGRAPHIC	**Regional or national**	Data at a regional and national level can be either single domain, *e.g.* national library, or cross-domain, *e.g.* collectionstrust.org.uk
	Pan-European	Represent a specific segment or sector of cultural heritage by aggregating data on a European level, *e.g.* TheEuropeanLibrary.org
	National Aggregation Initiatives	Organisation(s) appointed by their Ministry to take on the role as aggregator in the country, and can be cross- or domain-specific. A cross-domain example is Hispana.mcu.es

Source: Europeana. Aggregators and providers.

Accessibility

An aggregator may or may not have a portal where the content is made accessible to the public. For instance, this is the case of the Italian national aggregator CulturaItalia.

If the aggregator's portal is not accessible to the public, it is referred to as a dark portal or a dark aggregator. For instance, this is the case of MINT, the aggregator used within the Linked Heritage project.

Europeana Network

> *"The Europeana Network is an open forum for experts across Europe, from content providers and aggregators to providers of technical, legal and strategic knowledge and the creative industries. This forum represents and reflects the*

diversity of Europeana and gives its members an opportunity to voice their views.
[...]
Every individual that contributes data, technology or knowledge to Europeana is encouraged to join the Europeana Network. Thus, if you have a stake in Europeana, then please register for membership".
[See: Europeana Network].

Top

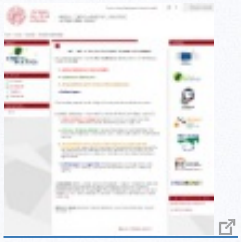
« Back Forward »

Linked Heritage LEARNINGObjects

MINT SERVICES

2/6 – THE LINKED HERITAGE PROJECT AGGREGATOR

Readings and training materials



Page Index

Linked Heritage (2011-2013) is an initiative coordinated by the Central Institute for Union Catalogue of the Italian Libraries, depending on MiBAC that extends and implements the ATHENA results. It is a best practice network funded within FP7 that began in April 2011 and will run for 30 months; it will contribute new content to Europeana, from both public and private sectors (mainly publishers), improve the quality of content in terms of richness of metadata, potential reuse and uniqueness, explore the potential of cultural Linked Open Data, and enable better search, retrieval and use of the content published in Europeana.

The Linked Heritage content providers

Twenty-two countries are members of the Linked Heritage consortium: culture ministries, government agencies, museums, libraries, and national aggregators, major research centres, publishers and small businesses, as well as organisations that contribute to Europeana for the first time with 3 million records of various cultural content.

Top

Aggregating content

Metadata Interoperability Services (MINT)

MINT: Metadata Interoperability Services compose a web-based platform that was designed and developed to facilitate aggregation initiatives for cultural heritage content and metadata in Europe.

It functions as a server for content ingestion and is based on open source software developed by the National Technical University of Athens (NTUA) in the context of the ATHENA project.

MINT allows content providers to upload, map, validate and deliver metadata to be sent to Europeana in an only web environment.

The platform also provides a management system both for users and organisations that allows the deployment and operation of different aggregation schemas with corresponding user roles and access rights.

Top

Mapping content

Metadata records are critical to the documentation and maintenance of interrelationships between information resources, and are being used to find, gather, and maintain resources over long periods of time. Descriptive, administrative, technical, and preservation metadata contribute to the management of information resources and help to ensure their intellectual integrity both now and in the future.

A crosswalk provides a mapping of metadata elements from one metadata schema to another. Supporting the ability to retrieve the same or similar content in different data source, crosswalks support the so called semantic interoperability.

The Linked Heritage Technology Platform, MINT, implements an aggregation infrastructure offering a crosswalk mechanism to support the following critical activities:

- harvesting and aggregating metadata records whether by standard and proprietary schemas
- migrating from content providers' model to a reference model
- transforming records from the Linked Heritage model to the Europeana Semantic Elements and the Europeana Data Model.

LIDO as Linked Heritage metadata reference model

MINT allows mapping and transformation of metadata into LIDO records.

LIDO stands for Lightweight Information Describing Objects. It is the result of a collaborative effort of international stakeholders in the museum sector, starting in 2008, to create a common solution for contributing cultural heritage content to web applications.

LIDO is based on CIDOC-CRM conceptual reference model. It comes from the integration between CDWA Lite and museumdat metadata schemas and it is based on SPECTRUM standard. Being an application of the CIDOC-CRM, it provides an explicit format to deliver (museum's) object information in a standardised way.

MINT implemented LIDO as intermediate harvesting schema. Initially conceived for museum sector needs, it is currently used in cross-domain contexts proving its adaptability and effectiveness in preserving the integrity of rich metadata.

The ESE metadata profile

MINT allows to convert LIDO metadata records into Europeana Semantic Elements (ESE).

ESE is a data model based on a Dublin Core-based set of fields with additionally 12 specific Europeana elements. Content providers must conform their metadata to ESE profile necessary for records to display correctly in Europeana.

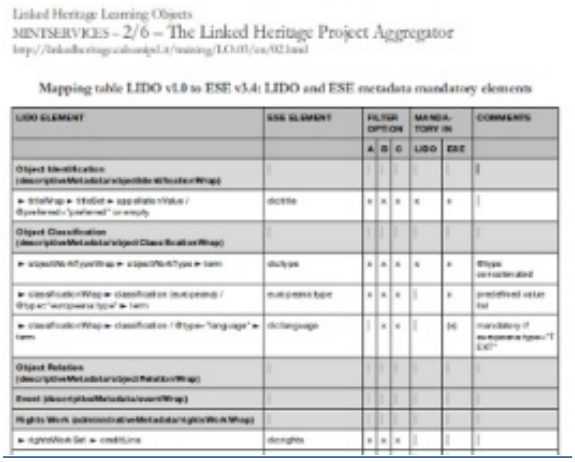
ESE is a subset of the Europeana Data Model (EDM), the new application profile that will be implemented in the coming months. Even EDM profile is supported by MINT (see: MINT screencast EDM Ingestion Tool).

Mandatory metadata elements

The ESE and LIDO metadata mandatory elements are the following:

- dc:title
- dc:type
- europeana:type
- dc:language (mandatory if europeana:type="TEXT")
- dc:identifier
- europeana:dataProvider
- dc:source
- europeana:isShownAt
- europeana:object
- europeana:isShownBy
- europeana:rights

- dc:rights


Linked Heritage Learning Objects
MINTSERVICES – 2/6 – The Linked Heritage Project Aggregator

Mapping table LIDO v1.0 to ESE v3.4: LIDO and ESE metadata mandatory elements

Click to download (PDF)

The complete mapping table LIDO v1.0 to ESE v3.4 by Regine Stein (Philipps-Universitaet Marburg - Bildarchiv Foto Marburg) is available in *Use of Content in Linked Heritage and Europeana (v.5), Annex 3*, prepared by the Linked Heritage DEA Task Force (see also *Content aggregation: tools & guidelines*).

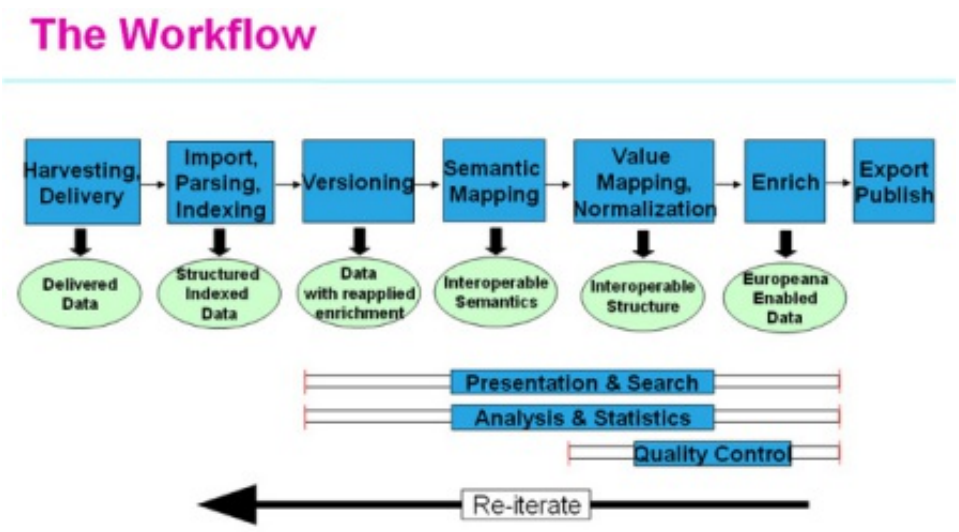
Metadata flow

MINT functions as metadata ingestion server, enabling content providers:

- to upload their datasets, that can be structured in heterogeneous metadata schemas, and map them to LIDO
- to transform metadata records into LIDO records and convert them into ESE
- to validate content through the Europeana Content Checker
- and to transmit content to the Europeana ingestion office via OAI-PMH Protocol

However the content providers can be asked by Europeana to check again the quality of the content once published online and to assess possible problems.

The graphic below summarises the metadata ingestion flow in MINT and the metadata flow towards Europeana:



The Workflow. Linked Heritage, June 2013

Top

Licensing content

Europeana Data Exchange Agreement

Europeana Data Exchange Agreement is the new licence adopted by Europeana in September 2011. DEA foresees that descriptive metadata (not the thumbnails) are subjected to the Creative Commons CC0 1.0 Universal Public Domain Dedication, which effectively means releasing content as public domain and allowing the commercial reuse of metadata.

This implies the possibility for Europeana to support open re-use of data and to publish metadata as Linked Open Data (LOD).

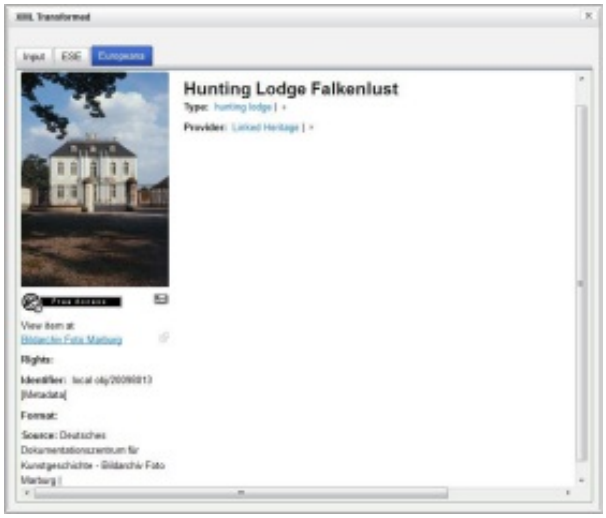
The DEA subscription is mandatory for all content providers to make available their collections in Europeana.

The Linked Heritage DEA Task Force

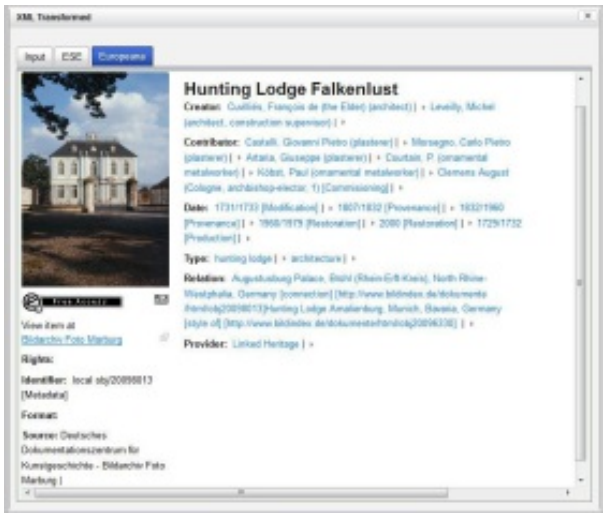
The Linked Heritage DEA Task Force was set up in order to present the Linked Heritage consortium with practical ways to fulfil the project duties (while implies a DEA subscription) and to keep the integrity of their data.

The task force elaborated a strategy that gives content providers 3 options for the metadata publication:

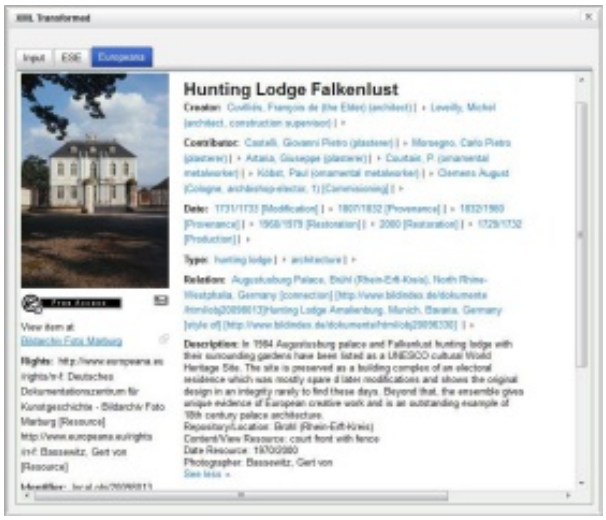
A. Publish a minimal metadata set to Europeana: of the metadata that is supplied to the Linked Heritage ingestion tool by the Content Provider, only the LIDO & ESE mandatory elements will be transmitted to Europeana under the Creative Commons CC0 1.0 Universal Public Domain Dedication



B. Publish an intermediate metadata set to Europeana: of the metadata that is supplied to the Linked Heritage ingestion tool by the Content Provider, all metadata elements will be transmitted to Europeana under the Creative Commons CC0 1.0 Universal Public Domain Dedication, except the LIDO elements that result in dc:description. This means that no object description, the part that most likely contains sensitive or valuable content, will be shown on Europeana



C. Publish a full metadata set to Europeana: of the metadata that is supplied to the Linked Heritage ingestion tool by the Content Provider, all metadata elements will be transmitted to Europeana under the Creative Commons CC0 1.0 Universal Public Domain Dedication.



MINT implemented a filter option to enable content providers to select the favourite one during the aggregation process (see screencast How to set a metadata filter in MINT).

Despite the minimum set of mandatory metadata required, Europeana as well as Linked Heritage encourage content providers to publish the widest range of information that can be made available through the Europeana portal, both for a better exposition and exploitation of content by end users and to make the user experience richer.

At present most Linked Heritage partners subscribed to DEA.

Top

Providing content

The Linked Heritage methodology: workflow

Assessing the Linked Heritage content providers' digital collections (the Linked Heritage survey)
The first step for taking the content into Europeana is assessing the digital collections that content providers described in the Description of Work (PDF), available in the Reserved Area of the Linked Heritage site.

This assessement can be easily done, for example, through a template. Linked Heritage content providers were asked to answer to a survey providing the following information:

- Country
- Data provider
- Primary contact
- Technical contact
- Collection URL
- Amount of metadata to be aggregate
- Amount of digital objects linked to metadata
- Object types: image, text, sound, video
- Description
- Metadata formats used
- Rights

As Europeana aggregates only metadata, it is of paramount importance to ask the amount of metadata and the amount of digital objects separately because the ratio *1 metadata : 1 digital object* can not always be the rule.

Training and training materials

Face-to-face training sessions with the project content providers were organised to train them on LIDO mapping and MINT use.

After the training workshop documents were delivered to all project partners.

Moreover, a specific section devoted to *tools* and *guidelines* for content aggregation was contextually published within the Linked Heritage web site.

The Help-desk service and Frequently Asked Questions

A help-desk service was set up at the beginning of the project to support the content providers problems. FAQ were also elaborated and posted on the Linked Heritage Web site.

Community

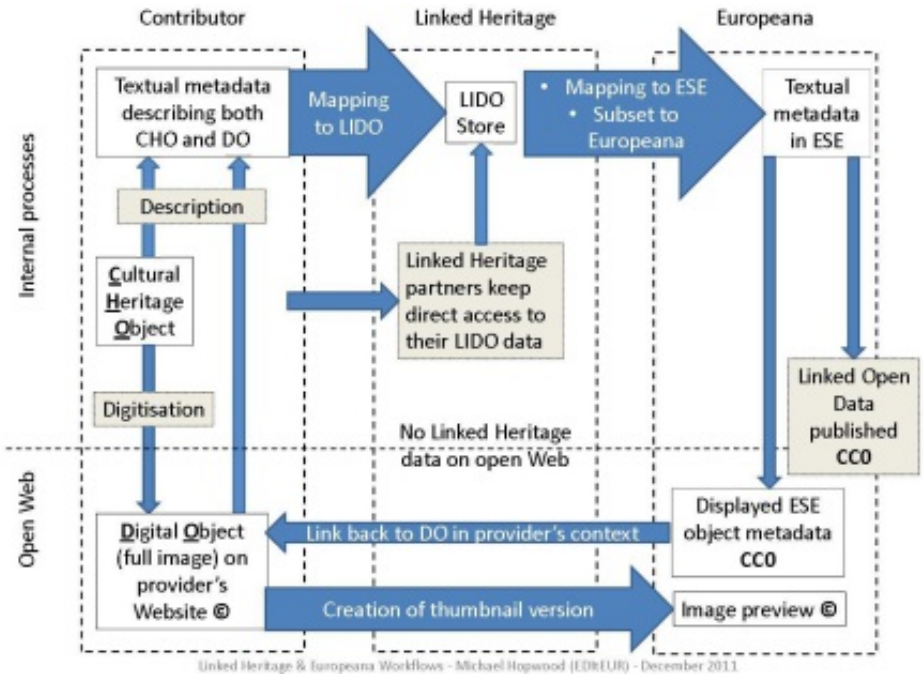
Workflow and feedback methodology is fundamental to assist content providers and keep the aggregation process under control; it also helps to build a sense of community.

Periodical interviews, constant review of the main aggregation issues, the analysis of data reports from MINT together with the ongoing updating of training materials are crucial tasks to consider for the benefit and the overall success of the project.

Top

Linked Heritage & Europeana workflows

The figure below summarises the way that metadata are contributed to Europeana through the Linked Heritage project.



Linked Heritage & Europeana Workflows. Michael Hopwood (EDItEUR), December 2011

Top

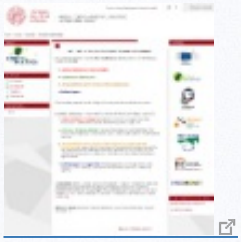
« Back Forward »

Linked Heritage LEARNINGObjects

MINT SERVICES

3/6 – MINT PLATFORM

Readings and training materials



MINT (Metadata Interoperability) Services, is the Linked Heritage technological platform that enables the content providers to:

- Upload metadata records, in XML or CSV serialization, from your computer or using the HTTP, FTP and OAI-PMH protocols
- Provide metadata records in a range of source formats (Dublin Core, MARC family, PICO ...)
- Convert metadata to LIDO, the Linked Heritage metadata reference model
- Submit the records to Europeana via the Linked Heritage gateway
- Manage data reports and administration functions

MINT is available at http://mint-projects.image.ntua.gr/linkedheritage/

Ingestion workflow

1. Register your organisation and login
2. Import your data
3. Create the appropriate mappings
4. Validate the metadata mappings
5. Convert metadata records into LIDO records
6. Perform the transformation to ESE
7. Enrich metadata
8. Deliver metadata to Europeana (ESE/EDM compliant)

COMING SOON

In next page, the screencasts series "Mapping workflow in MINT" will show you – step by step – the metadata ingestion and mapping workflow in the MINT platform.

Features

- MINT uses a visual mapping editor for the XSL language
- Mapping is performed through drag-and-drop and input operations which are translated to the corresponding code
- The editor visualises the input and target XSDs, providing access and navigation of the structure and data of the input schema, and the structure, documentation and restrictions of the target one
- It supports string manipulation functions for input elements
- Additionally, structural element mappings are allowed, as well as constant or controlled value assignment, conditional mappings and value mappings between input and target value lists

- Mappings can be applied to ingested records, edited, downloaded and shared as templates.

Preview interfaces

Preview interfaces present the steps of the aggregation such as the current input xml record, the XSLT code of mappings, the transformed record in the target schema, subsequent transformations from the target schema to other models of interest (*e.g.* Europeana's metadata schema), and available html renderings of each xml record. Users can transform their selected collections using complete and validated mappings in order to publish them in available target schemas for the required aggregation and remediation steps.

MINT Screencasts

EDM Ingestion Tool

This video demonstrates the EDM ingestion tool.

DPLA-Beta Sprint Demo

This video demonstrates the MINT platform and how it can be used in DPLA.

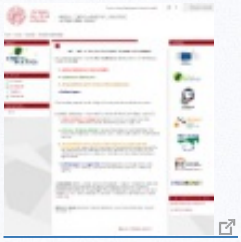
« Back Forward »

Linked Heritage LEARNINGObjects

MINT SERVICES

4/6 – PRACTICAL TIPS

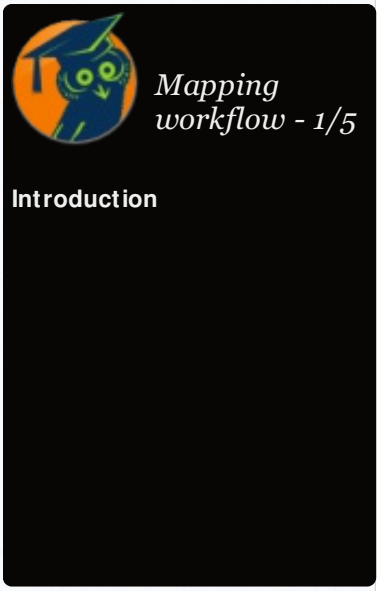
Readings and training materials



Page Index

The purpose of this section is twofold: first, to show the ingestion and mapping workflow in MINT, and second, to illustrate how to map metadata and digital objects rights. Each step is supported by a pertinent screencast.

Mapping workflow in MINT



Top

Mapping workflow - 2/5

Metadata Import

Top

Mapping workflow - 3/5

Define Root Item

Top

Mapping workflow - 4/5

New Mapping

Top

Mapping workflow - 5/5

Publish

EXPLORE FURTHER

For further documentation about the MINT platform, see MINT User Manual in MINT Wiki.

Top

Stating rights in MINT

Content providers and aggregators provide Europeana with metadata, a thumbnail and a link pointing to a digital object on the provider's Web site.

The terms regarding data delivery and re-use are governed by Europeana Data Exchange Agreement, that regulates Europeana's relationship with the content providers and aggregators, and sets out conditions for data re-use.

DEA is the central element of the Europeana Licensing Framework. The DEA sets out two simple principles:

1. For all **descriptive metadata** delivered to Europeana, content providers grant Europeana the right to publish the metadata under the terms of the Creative Commons CC0 1.0 Universal Public Domain Dedication,

2. Each **digital object** (and the associated preview) available via Europeana needs to carry a rights label that describes its copyright status. Content providers grant Europeana the right to publish the supplied previews. Previews may not be re-used by third parties unless the rights label related to the object allows such re-use

Top

Metadata rights

All the provided metadata will be published by Europeana as open data under the terms of the Creative Commons CC0 1.0 Universal Public Domain Dedication.

The Europeana Data Model contains a number of elements that are mandatory. Content providers are free to decide how much metadata to deliver, provided that

mandatory elements are given.
The Linked Heritage DEA Task Force elaborated a strategy that gives content providers three options for metadata publication in MINT interface:

- CC0
- CC0 (no description)
- CC0 (mandatory only)

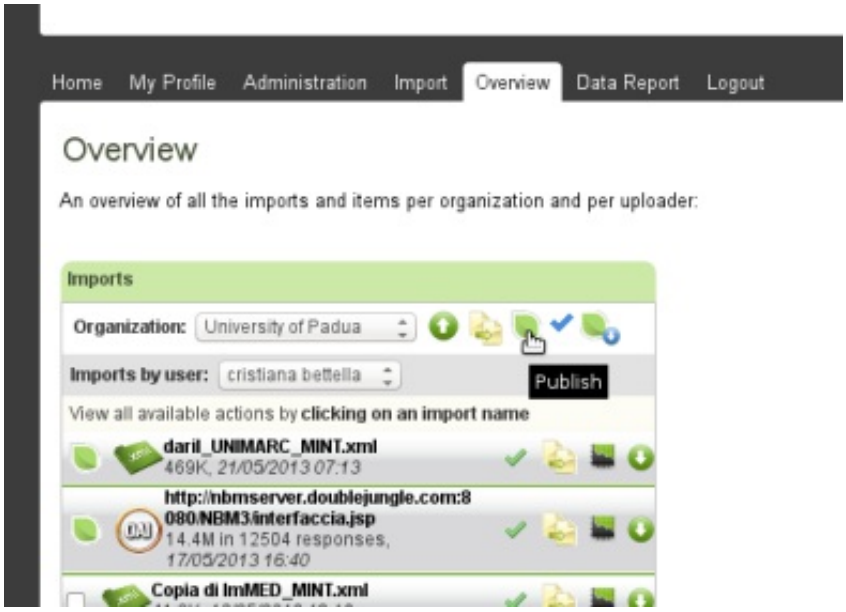
In order to map the filter

- open your mapping template in MINT, then select the RECORD wrap
- in *lido:**recordRights (europeana)*** » *lido:**rightsType*** » *lido:**term*** click on “enumerated” and select from the list the option to apply

The following screencast shows this procedure in MINT.

Download video in HD

The chosen option has to be confirmed when you publish the collection in MINT. At this point you can also decide to change the kind of filter to be applied.



Linking to the record source

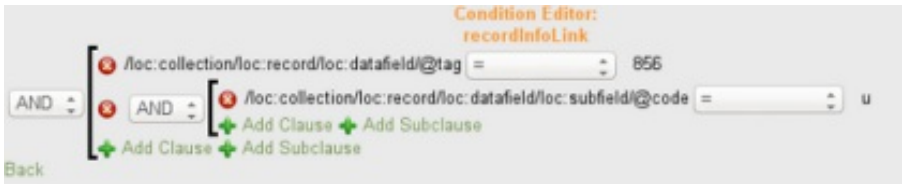
One of the mandatory elements is the link to the record source (in ESE is the element `europeana:isShownAt`).

To create the link, follow the directions shown in screencast "Linking to the record source in MINT", as summarised below.

This example refers to a mapping where the source schema is UNIMARC. The digital object URL is set in the field 856$u.

In the RECORD wrap open *lido:**recordInfoSet*** » *lido:**recordInfoLink***, then:

- from the source schema, drag “subfield” into “unmapped”
- click the grey star to set a condition, then from the source schema drag datafield@tag into “If condition”, leave the drop-down menu value as it is (“=”), and enter “856” in the field “Condition value”
- click the round grey icon to set “Complex condition”, then from the source schema drag “subfield@code” over the first occurrence of the word “undefined”. Leave the drop-down menu value as it is (“=”); double click on the second occurrence of the word “undefined” to open a box to set a constant value by entering “u”



The following screencast shows the mapping procedure described above.

Rights - 2/5

Linking to the record source in MINT

Download video in HD

Top

Digital object rights

Europeana's *Data Exchange Agreement* requires that content providers apply a rights statement for all digital objects described in their metadata.

The digital object rights are stored in the `europeana:rights` field, a mandatory element of the ESE profile.

The rights applied to digital objects will be inherited by the previews displayed through Europeana (*i.e* previews may be used by end users under the terms provided by the data provider through the `europeana:rights` value).

Only one rights statement can be assigned per resource. Rights statements are encoded as URLs referring to Web pages that contain information about the applicable rights. These Web pages inform users under which terms digital objects and their previews can be used.

Alongside each preview, Europeana will display a rights statement icon. By clicking on the icon users will be taken to a Web page explaining the statement.

EXPLORE FURTHER

Europeana Available Rights Statements.

How to set digital object rights in MINT

To map `europeana:rights` in MINT, follow the directions listed below:

- open your mapping template in MINT, then select the RESOURCE wrap
- open in sequence *lido:**rightsResource (europeana)*** » *lido:**rightsType (europeana)*** » *lido:**term (europeana)***
- in *lido:**term*** select from the list the rights statement connected to your collection
- to display the Europeana's rights statement icon you must type “preferred” in *@lido:**pref:preferred***

The following screencast shows the mapping procedure described above.

Download video in HD

Content provider rights

Any additional information about intellectual property rights, access rights or license agreements for the digital object can be included in the free text element `dc:rights`. If possible the value of this element will refer to the name of the rights holder.
Note that the element should not just duplicate the value in `europeana:rights`.

To map `dc:rights` in MINT, follow the directions listed below:

- In the RESOURCE wrap open in sequence *lido:**rightsHolder*** » *lido:**legalbodyID*** » *lido:**legalbodyName*** » *lido:**appellationValue*** » *lido:**legalBodyWebink*** » *lido:**creditLine***
- In *lido:**legalbodyName*** » *lido:**appellationValue*** type the content provider name (*e.g.:* DArIL ...)
- Optionally, you can add the content provider URL in *lido:**legalBodyWeblink***
- In *lido:**creditLine*** insert a text describing the use conditions established by the rights holder (*e.g.*: For information about using images, please refer to ...)

The screencast below shows this procedure in MINT.

Download video in HD

How to set the preview of the digital object in MINT

A further ESE mandatory element is `europeana:isShownBy`, where you code the image URL of the digital object.

Content providers have to set the image URLs in compliance with Europeana Portal Image Policy Image.

The example below shows how to code an image URL in MINT.

Also in this case the source schema is UNIMARC. The URL linking to the digital object is recorded in field 856$u, that is http://www.daril.eu/GiaGia1E1733AEC, *where "GiaGia1E1733AEC" is the digital object ID*.

In MINT the digital object URL has to be edited in order to get an image URL with this syntax: "http://www.daril.eu/fileadmin/photo/GiaGia1E1733AEC-(P01).jpg".
So we have to isolate the **digital object ID**, and build the image URL by concatenating the following strings:

- http://www.daril.eu/fileadmin/photo/
- **digital object ID**
- -(P01).jpg

Let us see step by step how to map the image URL above in MINT.

- In the RESOURCE wrap open in sequence: *lido:**resourceSet(europeana)*** » *lido:**resourceRepresentation(master)*** » *lido:**linkResource(master)***
- in the field "unmapped" of the *lido:**linkResource(master)*** insert the first string of the URL, the one preceding the digital object ID (*e.g.:* http://www.daril.eu/fileadmin/photo/)
- click on plus icon ("Add a mapping placeholder for concatenation") to add the second condition. This condition provides directions to locate the identifier
- from the source schema, drag “subfield” into “unmapped”; click the grey star to set a condition, then from the source schema drag datafield@tag into “If condition”, leave the drop-down menu value as it is (“=”), and enter “856” in the field “Condition value”
- click the round grey icon to set “Complex condition”, then from the source schema drag “subfield@code” over the first occurrence of the word “undefined”. Leave the drop-down menu value as it is (“=”); double click on the second occurrence of the word “undefined” to open a box to set a constant value by entering “u”

- set the “Apply function” by clicking on "F" icon, select “substring after” from the list and type in the field called “Select part of string after” the segment of URL preceding the identifier (*e.g.*: http://www.daril.eu/). The identifier will be thus concatenated to the first segment of the URL previously set (*e.g.*: http://www.daril.eu/fileadmin/photo/GiaGia1E1733AEC)

- finally add the third condition by clicking on plus icon (“Add a mapping placeholder for concatenation”) and set the second string of the image URL (*e.g.*: -(P01).jpg)

The screencast below shows the mapping procedure in MINT.

Download video in HD

Top

« Back Forward »

Linked Heritage LEARNINGObjects

MINT SERVICES

5/6 – CONCLUSIONS

Thank you for viewing our Learning Object *Mint services*.

We invite you now to answer some questions for your learning assessment and to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: Mint services.

For further details:

Readings and training materials

« Back Forward »

Linked Heritage LEARNINGObjects

MINT SERVICES

6/6 – CREDITS






Project by:

Luisa Banzato
Cristiana Bettella
Elena Bianchi
Lisetta Dainese
Massimo Ferrante
Michele Visentin
Antonella Zane
Bruna Zanusso

CAB, University of Padua (Italy)

Marzia Piccininno
Ministero per i Beni e le Attività Culturali - ICCU (Italy)

Valentina Vassallo
The Cyprus Institute - STARC (Cyprus)

Athanasios (Nasos) Drosopoulos
Katerina Komninou
Kostas Pardalis
Nikolaos Simou
Kostas Sismanis
Eleni Tsalapati

NTUA - National Technical University of Athens (Greece)

Texts by Cristiana Bettella (CAB, University of Padua), Athanasios (Nasos) Drosopoulos (NTUA - National Technical University of Athens), Marzia Piccininno (Ministero per i Beni e le Attività Culturali - ICCU), Valentina Vassallo (The Cyprus Institute - STARC), Bruna Zanusso (CAB, University of Padua).

Screencasts:

CAB, University of Padua (Italy)
NTUA - National Technical University of Athens (Greece)

Graphic and Web design by Gianluca Drago
CAB, University of Padua (Italy)

Funded by the Competitiveness and Innovation Framework Programme, 2011.

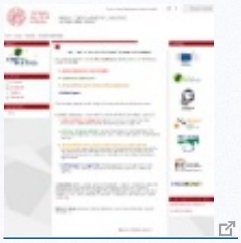
« Back

Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

ABOUT THE LEARNING OBJECT

Readings and training materials



Summary

This Learning Object is made up of two parts: one describes, through Europeana's eyes, the reasons why public cultural heritage institutions should contribute to Europeana and how. The other presents the University of Padova experience within the EU-funded project aggregator “Linked Heritage” such as formal and technical steps, timeline and other information useful for the project managers of content providers.

Target audience: decision makers and project managers of public libraries, museums, archives.

Commercial organisations willing to contribute to Europeana can refer to the Linked Heritage Learning Object Public-Private Partnership with Europeana.

Languages

- English

Forward »

Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

1/8 – EUROPEANA

Digitisation and online accessibility of cultural material are essential to highlight cultural and scientific heritage and to inspire the creation of new content. They help to democratise access to knowledge and to develop the information society.

Europeana is considered as one of the richest and authoritative sources of cultural information coming from European institutions. It is a portal that acts as an interface to millions books, journals, paintings, pictures, films, videos, audios, museum objects and archival records that have been digitised throughout Europe.

Endorsed by European Commission since 2005, Europeana's aim is to accumulate digitised content, standardise the metadata that describes it, apply linked data techniques to enrich it and promote persistent identifiers to locate it in the long-term.

Readings and training materials

Connecting institutions to enlighten cultural heritage.

Promo on why and how to contribute to Europeana realised in winter 2012-2013 by the University of Padova Library System in the context of a lab focused on multimedia resources editing.

Download video in HD

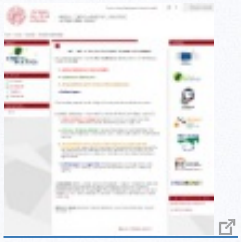
« Back Forward »

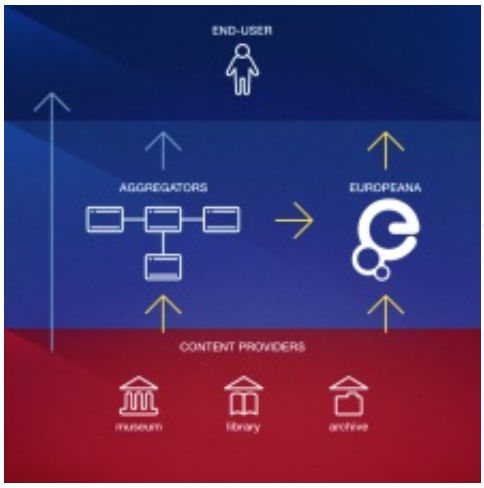
Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

2/8 – WHO PROVIDES DATA?

Readings and training materials





Europeana could not exist or continue to develop without the support of its many partners: thousands of organisations and institutions from all domains across Europe, from small cultural institutions to large national aggregators and from data providers to technology solutions.

An aggregator is an organization that collects metadata from a group of data providers and transmits it to Europeana. Three aggregation types are currently in Europeana:

- Country aggregators (regional or national)
- Project aggregators (within a specific theme or by domain)
- Independent organizations (with no Ministry or other mandate)

Even individuals are invited to collaborate to the project providing historical documentation (e.g. citizens wishing to contribute to Europeana special collections, *i.e. World War One Family History*)

Europeana encourages the contribution of aggregators that fit the needs of different countries, domains and users. Each aggregator, in turn, supports the data providers with administrative and technical operations, and training.

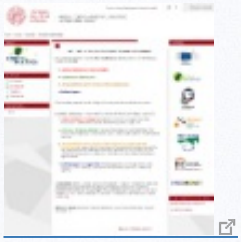
« Back Forward »

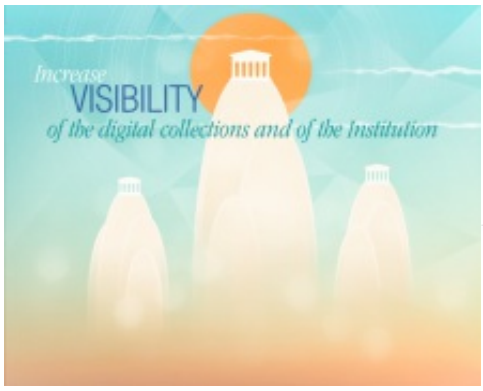
Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

3/8 – WHY PROVIDE DATA?

Readings and training materials



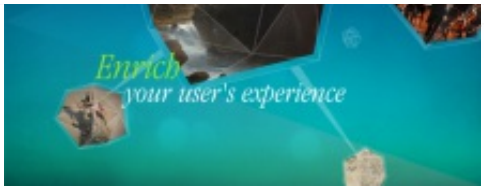


Thanks to the worldwide prestige achieved by Europeana and the importance attributed to the project not only by European Governments and the EC, but also by other international organisations and projects such as the Digital Public Library of America, institutions contributing data to Europeana easily gain visibility. [1]

The results of a Europeana survey highlight that users prefer visiting the websites of the content-providing institutions. They find it useful to view the digital object in its original context thus increasing traffic to the websites of the original institutions. [2]

*"Europeana adds value to the content by **juxtaposing** related images, texts, videos and audio items, thereby **'repatriating'** contents geographically dispersed into a single, coherent and contextual virtual space".* [2]

"Europeana enriches users' experience, giving them the opportunity to study related content, which comes <="" blockquote="">, and offers them in a single virtual space. Moreover, the multilingual



interface makes it possible for users to search in their native language and retrieve objects that are otherwise inaccessible to them". [2]

"Knowledge transfer is a key reason for being part of the Europeana network. ***There are critical issues that all European content providers and aggregators deal with****, including object modelling, semantic and technical interoperability of data, multilingual access, IPR (Intellectual Property Right) and business models for sustainability.* [2]

Even small institutions which would not otherwise have enough human, financial, and technical resources, may be part of Europeana. [2]

"Europeana will expose metadata to search engines, making deep web content accessible". [3]

Cost benefits

To participate in Europeana is an investment both from the cultural and economic point of view as well as an opportunity to have access to European calls for proposals and funding.

DIRECT BENEFITS **Benefits for Europeana and its participating investors and partner institutions**	• increased visibility of the contents • decreased cost of providing access through open source code • cost savings by standardization of metadata • cost savings through knowledge transfer
INDIRECT BENEFITS **Benefits that, through market transactions, are transmitted to consumers and producers in other markets**	• time saving by researchers • creation of new businesses in area's such as tourism • creation of new jobs in the creative industries • cost savings in the educational market

EXTERNAL BENEFITS **Benefits that can be seen as positive consequences of the actions of Europeana**	• the strengthening of a shared European Culture that help foster Euroepean identity • the promotion and enabling of diversity • the increase of social inclusion • the improvement of multimedia literacy

Source: Europeana - Cost & Benefits (PDF).

NOTES

1 The Europeana Aggregator Handbook 2010 (PDF) ↑
2 The Europeana Aggregator Handbook 2011 (PDF) ↑
3 Europeana Professional - Provide data ↑

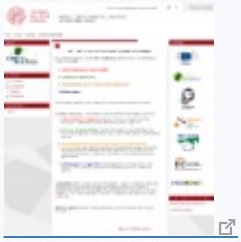
« Back Forward »

Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

4/8 – PROVIDE DATA: HOW AND WHAT

Readings and training materials



How to participate in Europeana

A partner institution interested in participating in Europeana will follow a procedure of subsequent submissions of information about its objects that will lead to the final approval of the request by Europeana Operations Team.

The Operation Team will then supply guidance and support for the harvesting and metadata mapping. Guidelines and technical requirements documents will also be given for the ingestion of the data as well as a Content Checker, a web tool useful for the validation of the mapping.

On publication in the European Portal, but also in case of update or ingestion of new data, the European Operation Team gives notice to the Cultural Heritage Organisation.

This procedure summarises the steps that aggregators, projects and content providers need to follow in order to contribute their data **directly** to Europeana.

Indirect content providers (*e.g.* those organisations contributing to Europeana through an aggregator or project) have to meet the requirements provided by the consortium in which they are involved.

What to provide to Europeana

Content providers can contribute to Europeana with the following types of items:

- **Texts**: books, letters, archival papers, dissertations, poems, newspaper articles, facsimiles, manuscripts and music scores
- **Images**: paintings, drawings, prints, photographs, pictures of museum objects, maps, graphic designs, plans and musical notation
- **Video**: films, news broadcasts and television programmes
- **Sound**: music and spoken word from cylinders, tapes, discs and radio broadcasts
- **3D**: virtual 3D representations of objects, architecture or places

Source: http://www.europeana.eu/portal/usingeuropeana_results.html

The categories Texts, Images, Video, Sound and 3D can be used by the users to filter search results in the Europeana portal.

EXPLORE FURTHER

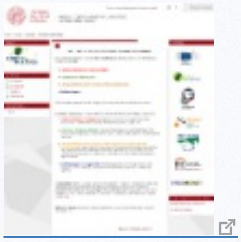
Europeana FAQs
Providers' FAQ
Support for Open Data FAQs

Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

5/8 – THE UNIVERSITY OF PADOVA EXPERIENCE

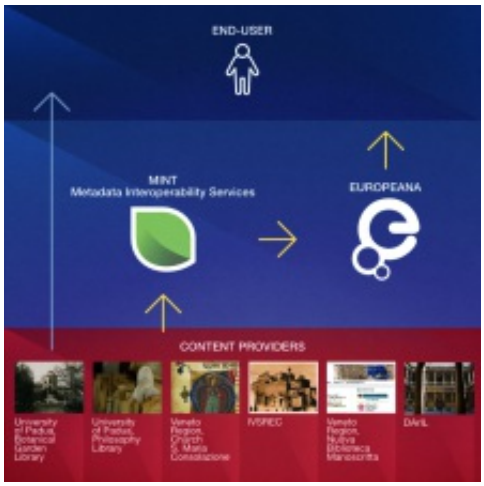
Readings and training materials



Page Index

This section describes the experience of an institution that provides data to Europeana through a Project Aggregator. In the context of Europeana, the term "project aggregator" refers to

> *"organisations that have joined a project consortium with a specific aim and purpose. Project aggregators aim to aggregate within a specific theme or by domain (single or cross)".*
>
> *Source: Europeana Professional – Aggregators and providers*



For further details see the *Aggregators* section in the Learning Object MINT Services.

The University of Padova took part in the CIP Best Practice Network "Linked Heritage" project which stemmed from the eContentPlus "Athena" project of April 2011 and ended in September 2013. Linked Heritage products and research results are available at www.linkedheritage.org.

Read more on the Linked Heritage Project Aggregator in the the Learning Object MINT : services.

Top

Why contribute to Europeana?

There are many reasons why the University of Padova decided to contribute to Europeana.

First of all, to make the cultural heritage of the University of Padova and the institution itself more visible through its digital collections.

In addition, we also thought it was important to contribute to Europeana through the European framework programmes and to obtain the organisational and financial support of EU-funded projects and initiatives.

This has prompted us to collect, organise and enrich metadata and other information on digitised contents which would otherwise be hidden in local databases or storage systems and therefore not available on the web.

Thus, adhering to Europeana has been valued not only as a short-, but also as a medium- and long-term investment from both cultural and economic reasons.

Another positive effect was that the Institution's librarians, curators, archivists and researchers could improve and share their expertise and knowledge about digital library systems and especially metadata standards, modelling and management, linked data, persistent identifiers, ingestion process, ontologies and digital copyright.

Our participation in Europeana, a prestigious initiative endorsed by the European Commission, attracted other local institutions. Consequently, new, valuable and unique collections have been added to Europeana thanks to the efforts and technical ability of our University staff.

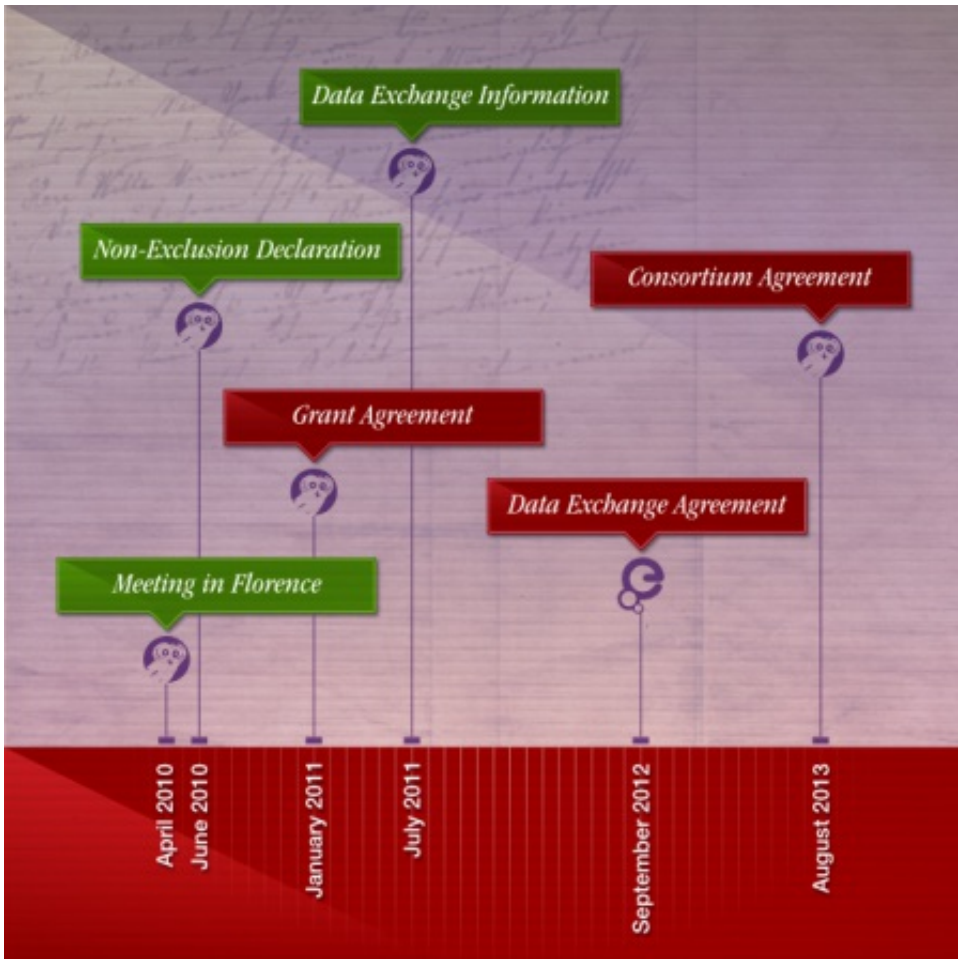
Top

How we provided data

Partnership formal steps

The project started on 23rd April 2010 when, during a meeting in Florence, the ICCU – Istituto Centrale per il Catalogo Unico proposed several cultural organisations to participate in a CIP Best Practice Network entitled "Linked Heritage" aimed at:

- providing large quantities of new content to Europeana, from both public and private sectors;
- enhancing the quality and richness of Europeana metadata;
- improving the functionalities available for search, retrieval and use of Europeana content;
- facilitating the dissemination and training on the key Digital Library concepts.



Timeline of the University of Padova contribution to Europeana through Linked Heritage

project.

Evaluation of the proposal

The University of Padova Library System decided to participate in Europeana through the Linked Heritage Project Aggregator as it considered it strategical to:

- cooperate with a community operating in the field of Digital Cultural Heritage, which includes representatives of all the key stakeholder groups from 22 countries: culture ministries, government agencies, museums, libraries, and national aggregators, major research centres, publishers and small businesses;
- participate in the discussion of the thematic working groups;
- be invited to conferences, international workshops and training courses organised by Linked Heritage;
- access the deliverables, technical reports and other materials produced by Linked Heritage partners within the scope of the project;
- use the Linked Heritage tools and technologies developed during the project;
- deliver cultural content to Europeana using the MINT platform implemented for the Linked Heritage project.

The decision: first formal act

The University of Padova Library System committed to participate in the Linked Heritage project by signing the "Non-Exclusion Declaration: Certification and Declaration on Honour" document on 1st June 2010.

The Grant Agreement

The ICT PSP Grant Agreement No 270905 between the Information Society and Media Directorate-General of EC and the ICCU (coordinator) and the Linked Heritage beneficiaries was signed in January 2011.
The Grant Agreement rules all provisions related to the contractual rules relating to financial contribution and implementation of the project.

The Consortium Agreement

The purpose of this Consortium Agreement is to specify with respect to the Project the relationship among the Parties, in particular concerning the organisation of the work between the Parties, the management of the Project, supplementary responsibilities of each partner not specified in the Grant Agreement and the rights and obligations of the Parties concerning inter alia liability, access rights and dispute resolution.
The Consortium Agreement was concluded between the Linked Heritage Parties on August 2013.

Agreements on data to be provided to Europeana

Data Exchange Information (the Linked Heritage survey)

In July 2011 the "Coordination of Content" (WP6) Leaders sent all partners a survey in order to assess the digital collections that content providers described in the Description of Work (PDF), available in the Reserved Area of the Linked Heritage site.

DESCRIPTION OF THE COLLECTION:
EXAMPLE TAKEN FROM THE "SURVEY OF THE LINKED HERITAGE PROJECT"
(AUGUST 2011)

Country	Italy
Provider	University of Padova
Primary contact	Name, Surname - email
Technical contact	Name, Surname - email
Collection URL	http://phaidra.cab.unipd.it/o:4714
Amount of metadata	2094

Amount of digital objects	3748
Object types	Images
Description	**Iconoteca dei Botanici**: portraits of Italian and foreigners botanists. The Botanists portrait collection includes 2,380 portraits of Italian and foreign botanists from the second half of the 17th century to the first half of the 20th century. The collection is made up of photographs (salted paper prints, albumen prints, aristotypes, platinum prints, gelatin silver prints), glass negatives, engravings, watercolours, drawings, paintings and photomechanical prints. The collection is preserved in the Padova Botanical Garden Library. Project description: http://www.bibliorto.cab.unipd.it/agenda/iconoteca-dei-botanici.
Metadata used	Dublin core; UNIMARC; University of Wien metadata; LOM
Rights	Public domain
Comments	Comments for rights: see terms and conditions available at: http://www.biblioorto.cab.unipd.it/in-biblioteca/dd-e-ill

Establish a partnership between Europeana and our cultural heritage organisations

The Europeana Data Exchange Agreement (DEA) between the Europeana Foundation and the University of Padova Library Centre was signed by the director of the University of Padova Library System on 8th August 2012 and by the director of the Europeana Foundation on 26th September 2012.

As of September 2011, the Europeana Data Exchange Agreement replaced all the existing agreements between Europeana and its data providers and aggregators, included the Linked Heritage project running since April 2011.

> *"With regard to licensing of the resources provided by data providers to Europeana, the [new] DEA sets out two simple principles:*
>
> 1. *For all descriptive metadata provided to Europeana, data providers grant Europeana the right to publish the metadata under the terms of the Creative Commons CC0 1.0 Universal Public Domain Dedication. This means that all metadata provided to Europeana can be re-used by third parties without any restrictions.*
>
> 2. *Each digital object (and the associated preview) that is available via Europeana needs to carry a rights label that describes its copyright status. Data providers grant Europeana the right to publish previews provided to Europeana. Previews may not be re-used by third parties unless the rights label related to the object allows such re-use."*
>
> *http://pro.europeana.eu/web/guest/data-exchange-agreement*

All the Linked Heritage content providers were asked to sign the new DEA before ingestion activities.

A dedicated task force was set up by the Linked Heritage coordinators in order to support each consortium participant to find out the proper solution. More information on the Linked Heritage DEA task force, as well as on the delivery modalities of metadata can be found in the Linked Heritage Learning Object MINT Services, *Licensing Content*

The University of Padova used the opt-out specification as offered in the DEA Art. 4 "Use of Previews", paragraph 3, in accordance with which Europeana will not be allowed to publish URLs pointing to the Previews together with other Metadata. Therefore, Europeana will only

use the Previews in accordance with paragraph 2 of the Art. 4: "Europeana is entitled to store and publish on Europeana.eu all Previews provided by the Data Provider, though only in combination with the Metadata that pertain to the same Content" (Art. 4, §2, §3, p. 4).

As the University of Padova Library System is in turn an aggregator of digital collections owned by local partners, we decided to discuss the main issues of the agreement with them in order to share awareness about the opportunities offered by the new DEA and the possible solutions.
IVSREC and the Region of Veneto signed the DEA directly with Europeana Foundation in May and in July 2013, respectively.

Technical steps

The ingestion workflow of the provided datasets into Europeana included the following steps:

- Ingestion activities: set up of the datasets for Europeana through the project aggregator technological platform MINT
- Harvesting of the provided datasets into Europeana
- Harvesting of the updated and new datasets into Europeana

A detailed description of the metadata ingestion and mapping workflow can be found in Mapping content and MINT Platform of the Learning Object MINT Services (See also the screencasts series Mapping workflow in MINT).

Top

Conclusion

The University of Padova Library System benefitted greatly from the Best Practice Network represented by the Linked Heritage Consortium. In fact, the whole ingestion process of the digital collections provided through the University of Padova was supported by our Linked Heritage partners who shared expertise and experience on digital content aggregation, as well as the scientific and technical coordination of the project.

In our experience, these are some of the advantages for public cultural institutions providing content to Europeana through project aggregators:

- economic returns due to:
  - increased visibility of the content of participating institutions;
  - improvement of the technical specialisation and know-how of their staff;
- significative time savings due to:
  - a coordinated approach to the ingestion activities with a common roadmap and deadlines scheduled in accordance with the aggregator coordinator;
  - accurate preliminary activity planning by each content provider. This step is essential as the status of source metadata (*i.e.* metadata completeness) may deeply affect project timing;
  - a prioritised channel which accelerate the ingestion process;
- significative cost savings due to:
  - the adoption of the open source software developed by the project aggregator (the ingestion tool and the terminology management platform)
  - the transmission of knowledge among Europeana content providers and aggregators
  - the opportunity for library, archive and museum working groups of the same institution to share methodology and collaborate in describing digital collections.

Top

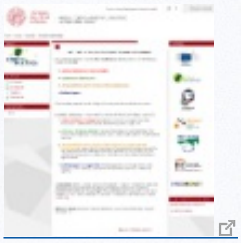
« Back Forward »

Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

6/8 – THE DIGITAL COLLECTIONS OF THE UNIVERSITY OF PADOVA AND ITS CONTENT PROVIDERS

Readings and training materials



Content Provider: University of Padova, Botanical Garden Library

The Botanists portrait collection

The Botanists portrait collection includes 2,380 portraits of Italian and foreign botanists from the second half of the 17th century to the first half of the 20th century. The collection is made up of photographs (salted paper prints, albumen prints, aristotypes, platinum prints, gelatin silver prints), glass negatives, engravings, watercolours, drawings, paintings and photomechanical prints.
The collection is preserved in the Botanical Garden Library.
https://phaidra.cab.unipd.it/detail_object/o:4714

Marie Rousseau, (190?).
Padova, University of Padova, Botanical Garden Library, IB.LL.41.

Nadar, *Ed. Boissier*, (187?).
Padova, University of Padova, Botanical Garden Library, IB.F.16.

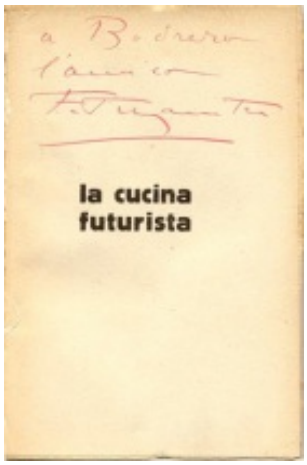
Content Provider: University of Padova, Philosophy Library

The handwritten dedications of the Bodrero collection

The collection includes 1,340 handwritten dedications generally addressed to the Italian philosopher Emilio Bodrero (1874-1949). The collection is preserved in the Philosophy Library of the University of Padova.

https://phaidra.cab.unipd.it/detail_object/o:8587

Philosophy Library, Emilio Bodrero collection.
Credits: Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia applicata.



F. T. Marinetti's handwritten dedication in *La cucina futurista*, 1932.
Padova, University of Padova, Philosophy Library, FB.E.863.

Content Provider: University of Padova, Library of Medicine “Vincenzo Pinali”

ImMed: Images in Medicine

The collection includes 800 images of historical and ancient documents held by the libraries of the School of Medicine of the University of Padova. The collection, represented by anatomical plates, photographs, slides etc., is considered of great historical interest.
https://phaidra.cab.unipd.it/detail_object/o:12473

Leopoldo Marco Antonio Caldani, *Tabula I: Exteriores humani corporis regiones exhibet in facie anteriori*, 1801.
Padova, University of Padova, Human Anatomy Library, K 1

Leopoldo Marco Antonio Caldani, *Tabula II: Regiones ostendit in superficie posteriori*, 1801.
Padova, University of Padova, Human Anatomy Library, K 1.

Content Provider: University of Padova, Libraries of Sciences, Pharmacology and Pharmaceutical Sciences

Scientific Wall Charts Collection

The collection comprises 179 images of wallcharts of the Library of Geosciences and of the Department of Biology of the University of Padova. The Biology wallcharts collection is stored at the Biological-medical library "A. Vallisneri". The collection is organised in 6 datasets as follows:

- *Wallcharts of the Department of Biology*: collection of 19 ascribed to Giovanni Canestrini (1835-1900), stored at the Biological-medical library "A. Vallisneri" https://phaidra.cab.unipd.it/detail_object/o:12794
- *Zoologische Wandtafeln herausg. v. Dr. R. Leuckart*: collection of 29 wallcharts of the Department of Biology stored at the Biological-medical library "A. Vallisneri" https://phaidra.cab.unipd.it/detail_object/o:12748
- *Dr. R. Leuckart und Dr. Nitsche: Zoologische Wandtafeln*: collection of 31 wallcharts of the Department of Biology stored at the Biological-medical library "A. Vallisneri". https://phaidra.cab.unipd.it/detail_object/o:12727
- *Zoologische Wandtafeln von Prof. Dr. Paul Pfurtscheller*: collection of 18 wallcharts of the Department of Biology stored at the Biological-medical library "A. Vallisneri" https://phaidra.cab.unipd.it/detail_object/o:12677
- *Collection Rémy Perrier & Cépède* of 16 wallcharts of the Department of Biology stored at the Biological-medical library "A. Vallisneri" https://phaidra.cab.unipd.it/detail_object/o:12656
- *Palaeontologische Wandtafeln*: collection of No. 64 wallcharts of the Library of Geosciences https://phaidra.cab.unipd.it/detail_object/o:12551

Mollusca. Cephalopoda. Tetrabranchiata, 1879-1901.
Padova, University of Padova, Geosciences Library, cassettiera TVP.

Copépodes parasites.
Padova, University of Padova, Biological-medical library "A. Vallisneri".

Content Provider: DArIL

DArIL: Digital Archive of Inaugural Lectures at Renaissance and Early Modern Universities

DArIL aims to facilitate scholars providing them with an access to a digital collection of searchable descriptions, digital photo-reproductions and codified transcriptions of inaugural lectures (also known as *paginae*). At present, DArIL comprises 110 documents preserved in a specific archive series in the Archivio Antico of the University of Padova: a series concerning inaugural lectures held by teachers of the artistae chairs at the Gymnasium Patavinum from the second half of the seventeenth century to the end of the eighteenth. Of these 110 documents, 24 are manuscript texts and 86 printed.
http://www.daril.eu


MATERIES ATQUE ORDO
PHYSICARUM DISCEPTATIONUM,
Q U A S
AN. MDCCLXVII-ET MDCCLXVIII
PER TRACTABIT
JOANNES DE ROMANIS
PHILOSOPHIÆ
IN PATAVINO ARCHI-LYCEO
PROFESSOR PUBLICUS.
EX TYPOGRAPHIA CONZATTI.
Superiorum Permissu.

Joannes de Romanis, *Materies atque ordo physicarum disceptationum quas pertractabit*, 1767.
Padova, University of Padova, Historical Archive, ARC.575.13

Content Provider: Veneto Institute for the History of the Resistance and the Contemporary Age (IVSREC)

IVSREC Photographs Collection

The photographs collection of the Veneto Institute for the History of Resistance and the Contemporary Age (IVSREC) comprises 919 images of episodes and events of the Resistance in northern Italy (867 photos, 30 Obituaries, 12 cards, 5 documents, 3 photocopies, 2 photomechanical reproductions and 1 slide).
The photographs were preserved in the Archive of the Institute. Some of them are from the Bundesarchiv in Koblenz.
http://phaidra.cab.unipd.it/o:12257

Solenni funerali dei caduti dell'insurrezione a Padova, 1945?.
IVSREC Archive, Foto: Scatola 6, n. 480.

Il Duomo di Padova dopo il bombardamento del 22-23/3/1944, 1944.
IVSREC Archive, Foto: Scatola 11, n. 672.

Content Provider: Veneto Region

Catalogue of the manuscripts in the libraries of Veneto: Nuova Biblioteca Manoscritta (NBM)

Nuova Biblioteca Manoscritta (NBM) is the online catalogue of the project of manuscripts cataloguing of Veneto libraries and of the software used for their description. NBM contains information about the libraries participating in the project and materials for manuscripts study and research. NBM is a project funded by the Veneto Region in collaboration with the Department of Classical Studies and Near Eastern Studies at the University Ca' Foscari of Venice.
http://www.nuovabibliotecamanoscritta.it/

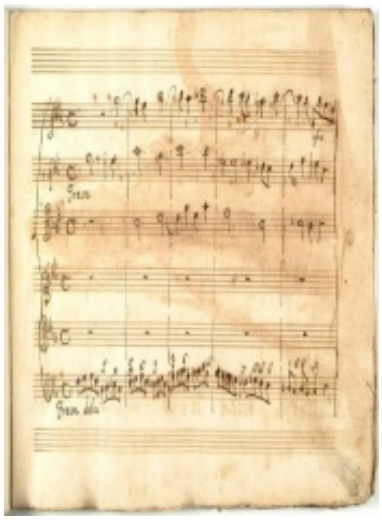
Padova, Biblioteca civica, B.P. 1099.
Tavola 1. PDBC_BP_ 1099_5v and Tavola 2. PDBC_BP_1099_5r.

Content Provider: Veneto Region

The music manuscript fund of the Church S. Maria della Consolazione in Venice, known as *della Fava*

The music manuscript fund of the Church S. Maria della Consolazione in Venice, known as della Fava, includes 72,531 images of about 764 works, 58 compositions in anthologies, and 896 sheets of fragments, with works, among the others, by Baldassare Galuppi, Niccolò Jommelli, and Johan Adolf Hasse.
The collection, currently guarded by the Redemptorist Fathers, is testimony of the intense activity of oratorios and sacred music representations, peculiar of the Congregation's mission of Filippini Fathers.
http://smcfava.regione.veneto.it

Cod. 72 (= Lit. 1) – Graduale, sec. XIV, primo quarto.
Capolettera figurato e abitato "G" (Gaudemus), c. CLVIr, S. Agata; motivi fitomorfi (132 x 84 mm). (Bottega di Neri da Rimini).
Venice, S. Maria della Consolazione Church, known as della Fava.



Giovanni Battista Pergolesi, *Stabat Mater*, prima metà 18 sec.
Venice, S. Maria della Consolazione Church, known as della Fava, Mus. Ms. 297.

« Back Forward »

Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

7/8 – CONCLUSIONS

Thank you for viewing our Learning Object *Why and How contribute to Europeana*.

We invite you now to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: Why and How contribute to Europeana.

For further details:

Readings and training materials

« Back Forward »

Linked Heritage LEARNINGObjects

WHY AND HOW TO CONTRIBUTE TO EUROPEANA

8/8 – CREDITS






Project by:
Luisa Banzato
Cristiana Bettella
Elena Bianchi
Lisetta Dainese
Massimo Ferrante
Michele Visentin
Antonella Zane
Bruna Zanusso

CAB, University of Padova (Italy)

Maria Teresa Natale
Ministero per i Beni e le Attività Culturali - ICCU (Italy)

Texts by Cristiana Bettella (CAB, University of Padova), Elena Bianchi (CAB, University of Padova), Lisa Dainese (CAB, University of Padova), Maria Teresa Natale (Ministero per i Beni e le Attività Culturali - ICCU), Antonella Zane (CAB, University of Padova).

Stopmotion animation by University of Padova Library System.
Music: Kevin MacLeod http://incompetech.com/
Sound: ERH http://www.freesound.org/people/ERH/
A special thank to Raffaella Traniello, multimedia projects consultant.

Graphic and Web design by Gianluca Drago
CAB, University of Padova (Italy)

Funded by the Competitiveness and Innovation Framework Programme, 2011.

« Back

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: COMMERCIAL AND HERITAGE VIEWS

ABOUT THE LEARNING OBJECT

Readings and training materials

Welcome to the series of case studies on persistent identification, presented by Linked Heritage partners led by EDItEUR. These studies show points of difference and similarity between the commercial media and cultural heritage sectors.

Intended audience:

- library and information professionals & masters students
- heritage (museum, archive, gallery) professionals & masters students

Learning outcomes – readers should aim to:

- understand what persistent identifiers are used for in the commercial sector (and a little of what they are *not* for!)
- know some of the "state of the art", latest developments in identifier technologies in the heritage sector
- start thinking about how digital cultural heritage online may be enriched in future with wider use of Web identifiers like DOI

Summary

The Linked Heritage project brought together information management standards and best practice from the cultural heritage sector with those of commercial media, such as

- book publishing
- recorded music
- film and tv
- photography

The aim was to investigate the potential for including commercial products in the Europeana Portal, adding the "gift shop" to the European Union's online museum, galleries, libraries and archives Web site.

To make this happen a team led by EDItEUR researched commercial and heritage information sharing methods. Our full reports can be found at the EDItEUR Web site for Linked Heritage.

This learning object and the four case studies it contains aim to share some of the knowledge gained, especially on the topic of persistent identifier systems – such as the ISBN or DOI - to enable library and information professionals, as well as museum, archive and other heritage professionals to engage with the best practice in the commercial world and see where common ground exists, with a view to enabling future public-private partnerships.

We present a viewpoint from the commercial sector to enable library and information professionals, as well as museum, archive and other heritage professionals to engage with the best practice in the commercial world and see where common ground exists, with a view to enabling future public-private partnerships.

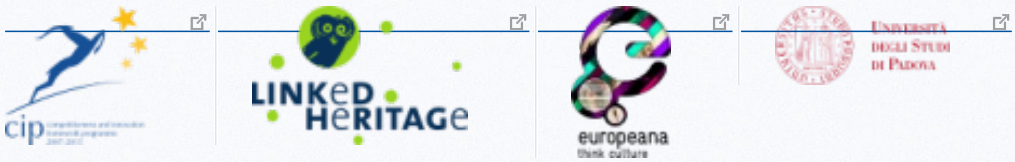
Contents

There are seven parts in this set – they can be read independently but fit together best in the order below:

Languages

- English

Forward »

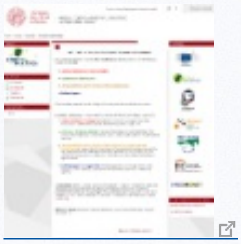


Linked Heritage
LEARNINGObjects

PERSISTENT IDENTIFIERS:
COMMERCIAL AND HERITAGE VIEWS

1/9 – IDENTIFIERS IN THE COMMERCIAL WORLD

Readings and training materials



Summary

This learning object aims to give you a background knowledge of the systems commercial media companies use to keep track of their books, music, films and photographs as they imagine, create and sell them to the public.

It will point out some things that are similar to work done in the heritage sector, and introduce some new ideas from the commercial world.

Intended audience:

- library and information professionals & masters students
- heritage (museum, archive, gallery) professionals & masters students

Learning outcomes – readers should aim to:

- understand what persistent identifiers are used for in the commercial sector (and a little of what they are *not* for!)

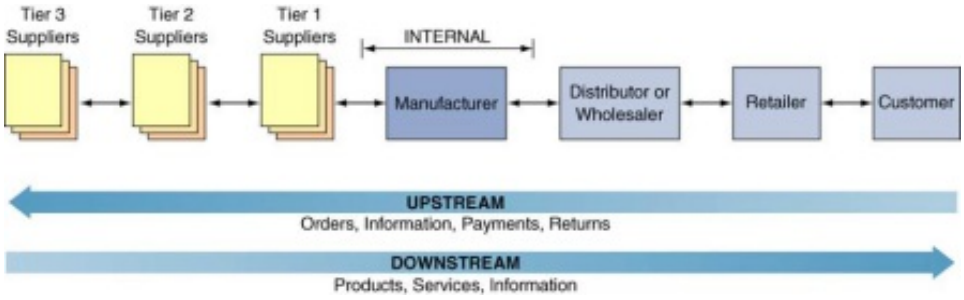
Page Index

How media companies handle information

A heritage organisation (a library, archive, gallery or museum) normally does as much of its activity "under one roof" as it can. It has many departments. Each specialises in one part of the "workflow" that gets a cultural object or work to the public.

In the modern commercial world, the story is very different. Of course, each media company has internal workflows. But it is very common that many different organisations each specialise in one aspect of the overall process that gets a product to its customer: *the "supply chain"*.

Below you can see a simple supply chain.



[1]

Of course, this is simpler than real life. In real supply chains:

- Many organisations work at each stage, sometimes cooperating, sometimes competing, in a "web" of partners;
- One organisation can sometimes fulfil many or all of these roles;
- Many problems can be found at each stage.

To make sure *the right information* flows each way in the supply chain, as you can see in the diagram above, *identifiers are used to make sure messages are understood correctly.*

An order must identify unambiguously which goods are being ordered. An invoice must identify which goods the requested payment is for. Identifiers - simple and unambiguous labels for the goods in question - make this more efficient.

The key is to understand that the identifier 'enables' something - that is, the identifier has a role and purpose. In this case, the purpose is to enable efficient trading of information up and down a supply chain. Ultimately, it is the nature of the processes that the identifier enables which defines the nature of the identifier itself. So for example, the well-known ISBN identifies a book product, because it is book products that are traded in the book supply chain. It matters critically whether the book is a hardback or a paperback, and whether it is a second or a third edition, because the customer specifically wants the third edition in paperback. So different ISBNs are used for hardback and paperback, even though they have the same text content, and new editions get new ISBNs, even though they might both be paperbacks and have the same title. The nature of the ISBN - and the rules used to allocate new ISBNs - are guided by what the ISBN is used for, delivering the right product to the customer.

Top

"Persistence" and why it matters

Now that museums and libraries present their collections on the World Wide Web, in a huge ocean of information, we are interested in using identifiers to help our "message" get through to the public, for example when creating "online exhibitions", or "digital libraries".

But the Web is a wild place and it takes time and effort to bring order to the chaos:

> *"[Identifiers] don't change: people change them."*[2]

In commercial terms, this is already a well known problem:

> *"Persistence is the consistent availability over time of useful information about a specified entity: ultimately guaranteed by social infrastructure (through policy) and assisted by technology"*[3]

In the supply chain above, note that *information* has to flow in both directions. This information may be:

- part of the products, maybe printed on them, or "embedded" in a digital file, or
- accompanying them, as a separate document, such as an order or an invoice.

Reliable, up-to-date information is crucial, so that

- legal contracts will be kept,
- commercial decisions will be effective,
- predictions and plans can be made, and
- inefficiencies can be avoided.

This is where *identifiers* are useful – they provide a clear, fixed point to attach all the relevant information as it travels along.

Identifiers are links...

...against which you can compile and later discover or retrieve richer information about the thing which is being identified. An identifier links the order to the exact goods being ordered, or an invoice to the goods that the payment covers.

In a complex supply chain with many products and types of product:

- there may simply be too much information to handle at any one point
- the products themselves may change;
- and the information about a product may change.

Links in the information chain can be *managed* to keep the latest, correct information available.

A unique identifier can bring you to the exact thing that you are looking for...

...or to further information about it, but having just one piece of information about it cannot be guaranteed to achieve that. A unique ID is like a key that opens up further possibilities.

"Persistence" in the commercial sector means "forever"...

...with clear contingency plans in place to address any risks that it might not!

In the commercial sector, the meaning of persistence is normally understood as "permanence" – that is, persistent for ever. For example, in the case of the published book supply chain,

> *"an ISBN identifies a given title and its edition and binding* ***for all time****."*[4]

Maintaining a commitment like this has its costs:

Identifers for (e-)commerce are a kind of "infrastucture investment" and thus are often administered by

- public sector bodies, or
- non-profit organisations working for an entire industry sector.

In many cultural heritage contexts, and many other public-sector organisations, "persistence" is defined in relative terms – relative, for example, to the lifetime of a project that creates the product to be identified, or the existence of the organisation that identifies the object, or perhaps relative to the expected usefulness of the resources identified.

Identifiers only identify things...

...they do *not* give any other legal, commercial or technical guarantees.

Media producers, wholesalers and retailers use identifier numbers and codes for one very clear purpose: to manage information about the things being identified. This information management process can continue throughout - and beyond - the life of the thing that's being identified.

Often people think that since identifier numbers are created using a formal process, and are publically available, there must be some kind of legal meaning or commercial guarantee attached to the product. There is not!

Of course it is possible for a public sector or commercial body to attach its *own meaning* to the use of identifiers from "outside" the supply chain. The technical design of the identifier system may even make it more efficient. But that is *not* their main purpose.

Top

EXPLORE FURTHER

Unique Identifiers: a brief introduction (PDF), by Brian Green and Mark Bide; 1999.
A short paper explaining the need for unique identifiers in the publishing business, discussing some of the basic issues raised in the Learning Object.

Top

NOTES

1 http://csbapp.uncw.edu/janickit/mis213/learning/module10/10-0.htm ↑
2 http://www.w3.org/Provider/Style/URI.html ↑
3 http://www.doi.org/doi_handbook/6_Policies.html#6.5 ↑
4 http://www.isbn-international.org/faqs ↑

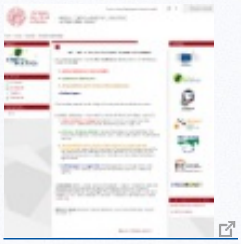
Top

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: COMMERCIAL AND HERITAGE VIEWS

2/9 – MISTAKEN IDENTITY: A FICTIONAL CASE STUDY

Readings and training materials



Summary

"Identifiers in the Commercial World" introduced the idea of a supply chain made up of organizations with different roles in creating and selling a media product. Product identifiers are the key to communication along that chain.

Here we offer a fictional and fun "case study" to help you understand some of the *practical ways* identifiers are *used* at each step of the supply chain. This should make the ideas in the real case studies clearer.

Intended audience:

- library and information professionals & masters students
- heritage (museum, archive, gallery) professionals & masters students

Learning outcomes – readers should aim to:

- understand *how* persistent identifiers work in practice in the commercial sector, and how that might apply in heritage work
- know *why* metadata has to be assigned to each identifier.

Page Index

Finding John Williams

Imagine you are a police officer, sent to find someone called **John Williams**. Your sources say he has stolen a large amount of money. You need to *identify* him and bring him in for questioning.

Files on two suspects have been brought to you. Can you *identify* John Williams from the two police artists' sketches and notes below?

His friends (who are mainly finance office workers) call him: "John Williams"...

...and *his* friends (who are mainly pirates) call him: "John Williams"...

You need to ask for more information on the "John Williams": do they want…

- Mr. John Williams, the finance office worker?
- or Captain John Williams, the pirate?

Top

What's in a namespace?

A person's *name* is a useful for finding that person – but only if there is only one *"John Williams"!* This is the problem of *uniqueness* and it depends on *context*.

Outside the contexts of the police investigation and the suspects friends, there could be thousands of people called "John Williams". *A personal name is not universally unique*.

The context in which a name or identifier for something or someone *is* unique is the "namespace" (OK, this is actually a very specific term in computer science, but the basic idea is similar).

Asking for John Williams by the same "name" in the contexts of the two sets of friends – the "namespaces" below – will get two different results. In the context of the investigation, that includes both groups, you can specify both the "local" name *and* the "space" where it applies.

Two "namespaces", one name...

Down at the docks, shouting "John Williams" will get you a pirate captain...

In the finance workers' bar, shouting the same name will get you the fraudster.

A phonebook might be an example of a "good namespace" in which you expect to find *one and only one* person referenced by each "name".

Each phone number is the "unique name" or *identifier* of that household or business. Note that it is the phone numbers that are unique names, not the names of the people. The phone numbers identify telephones that you may call, not people (there will often be more than one actual person with the same name, and one person may have more than one phone).

These examples hint at the fact that **any system of unique identification needs a human element, a social and organisational component to maintain it.**

This identifier *agency* must have a **public face**, to present it to the world at large and its users so they know which namespace they are dealing with.

Top

Making the right connections

You send officers to find both suspects. The two men might be out alone, so could not be recognised by their friends at the finance office or on the treasure ship.

What *identifying characteristics* could you give to your officers? Perhaps the file might look like this:

SUSPECTS CALLED "JOHN WILLIAMS"

Differentiating number	1	2
Gender	Male	Male
Typical facial expression	Angry	Angry
Hair colour	Black	Grey
Head wear	None	Pirate bandana
Eye wear	None	Eye patch (right eye)
Jewellery	None	Earring (left ear)
Facial hair	None	Beard and moustache

As your officer sees each of these characteristics, he becomes more certain that *this* is the *correct* "John Williams". These are the "metadata" for each suspect that enable the activity of identification. They are even more important for computer systems, which are unable to pick up other "clues".

But – for a case with *hundreds* of suspects, how could your officer *quickly and accurately* report which suspect they have seen?

Using a unique identifier!

This is a good idea, for the same reasons as we discussed in "Identifiers in the Commercial World". What would the identifier code look like?

A *structured* identifier for each person like "2013-case-1-name-j-williams-number-1" might be highly unique, and might make your filing system easier to organise.

Some problems might be:

- what if you need to *share* those files and their identifiers with police in other places, and you want to protect the privacy of suspects who, after all, might not be *the* John Williams you are looking for?
- what if some of the known details *change,* for example, the spelling of Williams' surname is found to be in fact "William" without the "s", when discovered in writing for the first time? Would referring to the suspect using an identifier built on the old information cause confusion? Or perhaps you might discover "William" to be a pseudonym?

An dumb number like the simple serial number above – with the namespace "Suspect number" – would be more appropriate:

- Extra "linking" details, like the year, the case number and the suspect's name, characteristics and differentiating number can be held in a separate, confidential file.
- They would be shared only with the appropriate agents, at the right time and place. Information that changes a lot, like spellings or even whole names, could then also be shared only after the latest update.

- Dumb numbers like these are also easier to transmit in various ways since they contain fewer details that can be mis-heard or mis-read.
- A number could include a little structure relating to the namespace *e.g.* part of the number might indicate the location of the police station where the file was first created.
- It could incorporate a check digit based on a defined mathematical formula which would help to recognise mistakes in transcription and transmission.

If a new officer takes over these files, using a "dumb" serial number to identify each suspect's file will avoid confusion.

- An identifier structured with details of when the file was created, by whom, and what the "linking information" was thought to be at the time. In this way the new officer would not be drawn into thinking that such details were still the latest news.

What if new details of a known, identified suspect come in? Maybe a new witness tells you that suspect 0000-0000-0001 actually has blond hair and glasses?

- There must be some *management processes* to keep other police services up to date with changes to the files, and perhaps *select between alternative reports* on the basis of how much you *trust* the information source;
- To *look up* these suspect numbers and *get accurate information*, agents on the ground will need to know *who maintains the files*, and have some mechanism for *contacting the authorised holder of the information* to retrieve it.

Commercial sector identifiers in real life use all of these features.

Top

Does all this apply to heritage organisations?

This fictional, highly exaggerated example, highlights some of the problems that identifier systems in the commercial and heritage sectors are expected to solve.

In the real world, cultural heritage professionals and information managers in the creative industries deal with *ambiguity and complexity on a very large scale*, especially in the Web environment.

For example, Wikipedia lists several "artists and entertainers" known as John Williams [1]

Full name(s)	Profession(s)	Earliest date	Latest date
John Williams	Stage, film, and television actor	1903	1983
John Williams	Classical guitarist	1941	
John Williams	Composer		
John Williams	Radio personality	1959	
John A. Williams	Novelist	1925	
John B. Williams	DJ	1977	
John David Williams	Musician and songwriter	1946	
John Edward Williams	Author of novels	1922	1994
John Ellis Williams	Novelist	1924	2008
John H. Williams	Film producer		

John Hartley Williams	Poet	1942	
John James Williams	Poet		
John McLaughlin Williams	Conductor		
John Richard Williams	Poet	1867	1924
Johnny Williams	Blues guitarist	1906	2006
Johnny Williams	Jazz drummer	1905	1984
J. Lloyd Williams	Botanist, author, and musician	1854	1945

In most of these cases, seeing the profession and vital dates of people with similar or the same name is enough to tell them apart; the larger database of ISNI (International Standard Name Identifier) lists 85 "creators" called John Williams, and in the Web environment of course the number of documents containing this common name may increase to thousands or millions.

Putting together a comprehensive online guide to "John Williams", or making a new exhibition or publication on "John Williams" stand out in this crowd will need a lot of work!

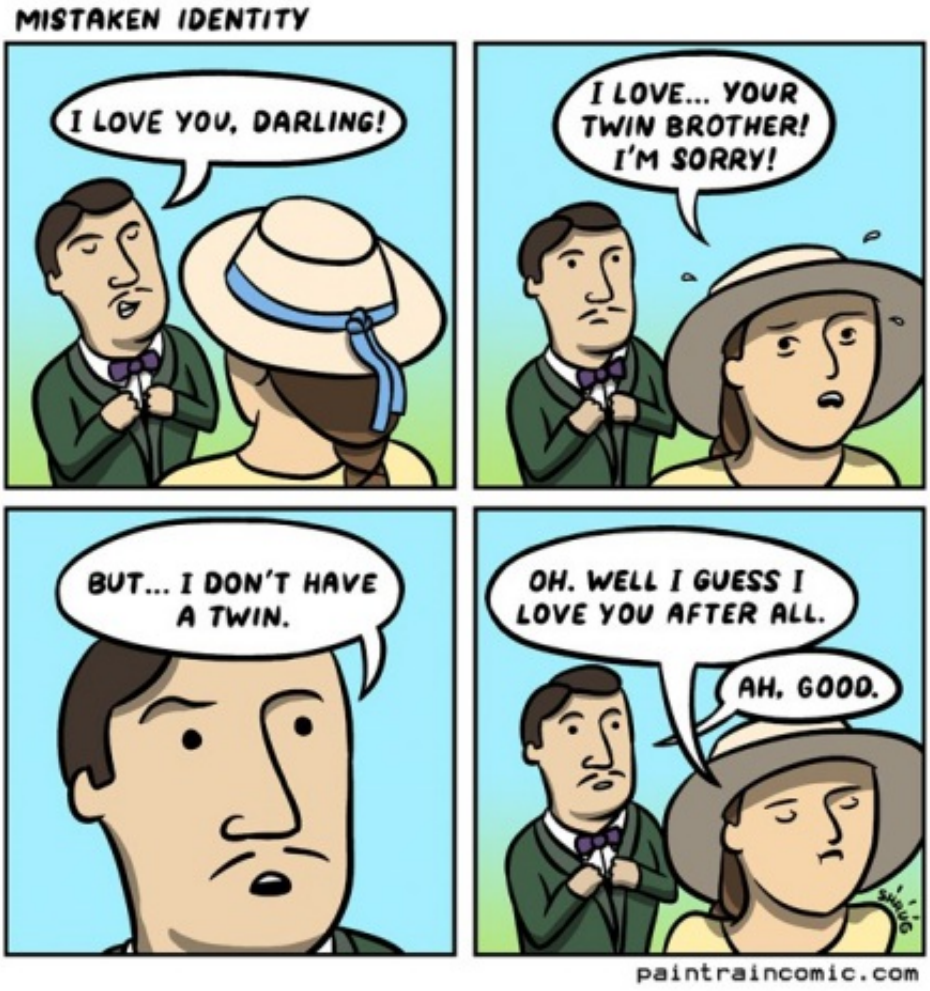
Top

Is everything relative?

A word needs to be said here about the meaning of "identity".

In the police story above, "identity" is a requirement for *similarity in every possible way*. There can be only one "right" criminal in the end. Jailing a John Williams who is "relatively similar" to the one who actually stole the money in question is not acceptable.

Many situations involving personal identity require absolute identity: *e.g.* curation of an exhibition on John Williams, or publication of one of the poet John Williams' poems. This is the type of "identity" we are most familiar with from daily life:



[2]

In the cultural heritage world, absolute identity is the normal goal of maintaining data and identifiers.

Visitors to a Pelé football museum expect to see *the* footballs used by Pelé, not just "a" football of a similar type, or perhaps *the* shirts worn by him during his World Cup winning matches, not just "a" shirt produced for his team.

- Heritage objects, even when many *similar* items exist, take on some of the "personality" of the events, things and people they are associated with:

  > *A museum object is more like **an illustration or witness** of the past, than information in its own right.* [3]

  Those unique *events* can only have happened to *one specific thing* – **museum objects are not substitutable**.

- In contrast, commercial products, especially media products that convey information, rather than "history", are intended to be only *relatively identical* – that is, they are *functionally* identical:

  > *What is "the same thing" ("a copy") for one user, purpose, or context will be "two different things" for another. The two users may have different purposes in mind when they ask "are X and Y the same?"; and as we have seen, this question is implicitly "are X and Y the same for the purpose of...?"* [4]

  For example, a request for "an English dictionary" *for the purpose of* learning the 1000 most common words for daily conversation may be satisfied by even very old editions. But to *for the purpose of* learning the latest jargon and buzzwords on the Internet, this "English dictionary" may not be identical after all:

A dictionary of "English" – and a dictionary of "the latest English"

CRYOTHERAPY	PENTREPRENEUR
SOCK PUPPET	ZERO-HOURS
DENCH	JIGSAW FAMILY
FACEKINI	ASTROTURFING
LINK JUICE	CONTACTLESS
SMACK TALK	PAY-NEUTRAL
EARTHQUAKE CLOAK	RE-MODE
FIDOCAM	MICROBIOME
MINI-JOB	NÉGOCIANT
MOOC	*etc…*

An English dictionary a few decades or even a century old might contain the 1000 most common words in daily use

A more recent (*i.e.* up-to-date) English dictionary might be needed if the 100 newest words are required as well.

5 6

We highly exaggerated the differences in user requirements for each English dictionary. But we are talking about at least two different products.

Commercial sector identification can be complex because *differences between two similar products become smaller and smaller, but still make a great difference to everyone involved,* from photographers, the authors of books and film scripts, and composers of music, to the eventual purchasers and readers, listeners and viewers.

The cultural heritage "view" focusses on "this item":

CULTURAL HERITAGE VIEW

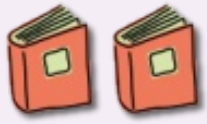
Particular item		Identification	Other "copies"
Pelé's World Cup winning football shirt		http://pele-world.museum/objects/00000000001 Unique interest; no substitutes possible	None possible. Each shirt considered unique due to its associations; even if otherwise very similar style, material etc.
John Hartley Williams' work shirt		http://j-hartley-williams.museum/objects/00000000001 Unique interest; no substitutes possible	
John James Williams' work shirt		http://j-j-williams.museum/objects/00000000001 Unique interest; no substitutes possible	

In the commercial world, the focus is on "this product" – which really refers to a *"set" or "class" of identical items. If there are three copies of a new book in the bookshop, it does not matter at all which of the three you buy - they are identical.*

The real value in media products is some unique *intellectual property.* Intellectual property by its nature *can* be reproduced, hence the importance of copyright law:

COMMERCIAL PRODUCTS VIEW

Particular product		Identification	Other copies
Old (but cost-effective) English dictionary		This product: ISBN 978…..1 (any copy is good as another)	 …probably limited number (it may be "out of print")

New English dictionary (published 2013)		This product: ISBN 978.....2 (any copy is good as another)	 ...probably possible to order more (in print) or even request new copies (print on demand) – or permanently available (e.g. as an ebook)
Not so new English dictionary (1997)		This product: ISBN 978.....3 (any copy is good as another)	 ...probably still available for sale but not necessarily "in print" or even available as an ebook; far less demand for slightly older dictionaries than for "the latest"!

Of course there are some *grey areas*, for example, museums of ephemera or popular culture, rare book dealers, or a publisher who deliberately issues limited editions of a fixed number of copies, and maybe other more unique features, like the author's signature [7]:

Top

EXPLORE FURTHER

Identifier Interoperability: A Report on Two Recent ISO Activities by Norman Paskin; DLib, 2006.
A technical paper explaining the importance of a minimum set of metadata attached to any identifier so that it can be used in practical applications.

Indecs metadata framework (PDF), by Godfrey Rust and Mark Bide; 2000.
A comprehensive technical guide to commercial sector metadata principles, including creation of identifiers.

Top

NOTES

1 http://en.wikipedia.org/wiki/John_Williams_%28disambiguation%29 ↑
2 Mark Pain © http://paintraincomic.com/mistaken-identity/ ↑
3 http://www.researchspace.org/researchspace-concepts/technological-choices-of-the-researchspace-project ↑

4 http://www.dlib.org/dlib/january03/paskin/01paskin.html ↑
5 http://32cherry.deviantart.com/art/book-264579890 ↑
6 http://dictionaryblog.cambridge.org/tag/neologisms/ and http://dictionaryblog.cambridge.org/tag/neologisms/page/2/ ↑
7 http://www.worldheadpress.com/special-editions-231 ↑

Top

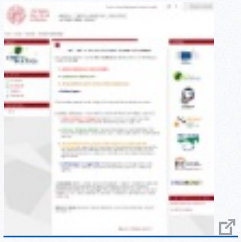
« Back Forward »

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: COMMERCIAL AND HERITAGE VIEWS

3/9 – CASE STUDY 1: THE ISBN, A CORE IDENTIFIER FOR PUBLISHERS, BOOKSELLERS AND LIBRARIES

Readings and training materials



Summary

This case study deals with an identifier system that is already familiar to many in the heritage sector, the ISBN (International Standard Book Number).

Librarians are familiar with ISBN numbers. They are used to

- order books for their collections
- add unique identification to their catalogue records
- retrieve books requested by their users
- collect reading list information in educational settings
- identify e-books that are “published” when new digital versions are created from older, out-of-copyright and out-of-print volumes.

Finally, many national libraries act as “registration agencies” for the ISBN standard in their countries. They gain access to information about new books directly from publishers this way, as well as copies of the books for cataloguing and legal deposit.
The National Library of Hungary, a Linked Heritage partner who contributed to this Learning Object, is one example of a national library that is also an ISBN registration agency.

Page Index

History and development

The ISBN is one of the success stories of international technical standardisation, having been adopted in the UK and USA in the late 1960s, and becoming an ISO standard just a few years later.

Now recognised in over 200 countries, it has increased in adoption and become integrated into other major standards, such as the GTIN-13 method for encoding retail product numbers in barcodes [1].

Top

Intended scope

An “International Standard Book Number” is clearly used for “books” in some sense, but a more precise definition is given by its official governing body:

> *“ISBNs are assigned to **text-based monographic publications** (i.e. one-off publications…) and **certain types of related products** that are available to the public, whether those publications and related products are available on a gratis basis or to purchase”*

Note that this is more to do with the type of “publication” (the method of *“making available to the public”*) than the *contents*, and that “available to the public” implies that there are intended to be enough of copies for everyone who wants one!

The ISBN does *not* just identify “monographs” in the sense of “relatively short book or treatise on a single subject”.
In “library cataloguing, **any nonserial publication**, complete in one volume” [2] is called a “monograph”.
The “nonserial publication” part is absolutely the same as in the ISBN system. But: a “volume” is a *bibliographic* idea, and not what the ISBN is necessarily intended for.
A “volume” may be *one specific copy* of a book, or a *conceptual division* of the text content *within the book*, certain not something an ISBN can indicate. Libraries normally give a local number to each specific copy they hold, and they do not use a new ISBN for each copy!

Some examples from the official ISBN F.A.Q. will make clearer what publishers and retailers refer to with ISBNs:

- Printed books and pamphlets
- Individual chapters or sections of a publication if these are made available separately
- Braille publications
- Publications that are not intended by the publisher to be updated regularly or continued indefinitely
- Individual articles or issues of a particular continuing resource (but not the continuing resource in its entirety)
- Maps
- Educational/instructional films, videos and transparencies
- Audiobooks on cassette, or CD, or DVD (talking books)
- Electronic publications either on physical carriers (such as machine-readable tapes, diskettes, or CD-ROMs) or on the Internet
- Digitised copies of print monographic publications
- Microform publications
- Educational or instructional software
- Mixed media publications (where the principal constituent is text-based)

The common factor here is that these publications are discrete, text-based products that can be obtained (for a price or for free) in a single, complete transaction; one transaction obtains access to the whole publication.

Top

Use cases

ISBNs are meant for the use of “publishers, booksellers, libraries and [internet] retailers” [3] to help with:

- ordering, supplying etc.
- listing, cataloguing, etc.
- sales records, invoicing and payment,

- stock control [4]

They are used by *intermediary partners* in the supply chain, such as

- books-in-print services who collect data about products and provide a service to those who need large amounts of reliable product information
- retailers wanting to offer a wide range of books to their customers
- libraries and library cataloguing agencies, who can save time and money by re-using some of the book data to create catalogue records

The ISBN is a perfect example of *supply chain identifier* as we described in the section "Identifiers in the commercial world". It is used by suppliers and purchasers at every point in the chain.

Top

A number of many parts

An ISBN is really simply a number made up of 13 digits. It was always meant to be easy to find *printed on the book itself*, and to copy down for human use as well as machine processing. Displayed on the product itself it acts as a kind of "dog tag" for the book "found in the wild"…

[5]

The number can be written with or without spaces or hyphens: **978-0-571-08989-5**, **9780571089895** and **978 0 571 08989 5** all mean exactly the same thing.

It is not obvious what the five parts mean – however, knowing the number is part of the ISBN namespace, it is possible to discover a certain amount of information within those digits.

There can be confusion about what an ISBN does and does not tell you, mostly because people assume *more* than is really there.
Essentially, it tells you a small amount of information about *where* and *by whom* the publication was registered.
So it links a publication to its publisher *at the time of registration* - but maybe not after that. The rights to sell that publication may change hands, or the publisher itself may change its name or go out of business. So, the information encoded in the ISBN itself is historical.

Let us take the real ISBN number **978-0-571-08989-5** as an example and describe what each section means:

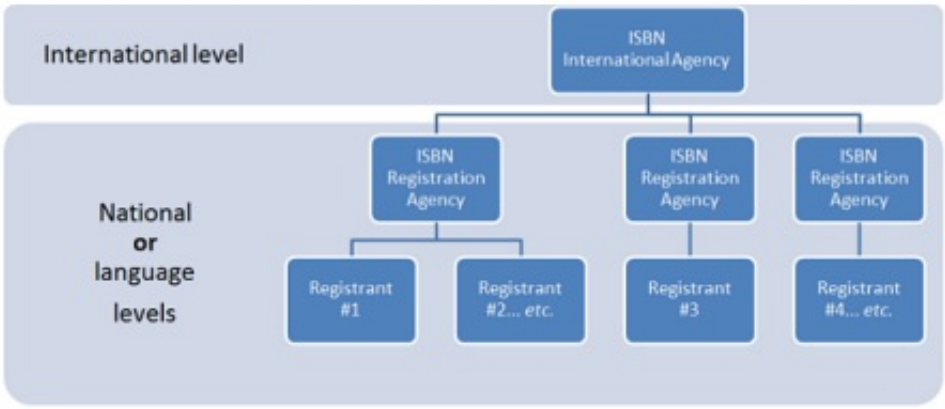
978	0	571	08989	5
• Prefix declaring this number as an ISBN within the GTIN-13 namespace • Could also be 979 (and more may be added in future to add more space)	• Registration group • Groups are either geographic or language-based • In this case "0" indicates English language	• Registrant • Identifies the publisher or imprint (brand name) which registered the publication	• Publication • Each specific edition of each publication gets a new number: e.g. hardhack, paperback, revised edition, different ebook formats	• Check digit • calculated from the other digits and used to detect transcription errors

Each part of this number *makes a link* between things and organisations.

This is the organisational and social infrastructure that gives the number both uniqueness and persistence.

There is a fairly simple hierarchical structure of agencies and registrants to ensure:

- firstly, uniqueness of assignment (authorisation to use ranges of numbers flows down from the International Agency, through RAs, to registrants)
- secondly, flow of product information (back up from registrants to the RA)



Because Registration Agencies can work on a national **or** language level, it is possible that there will be

- *more than one agency assigning numbers from groups for the same language* – such as the separate English-language agencies in the UK and USA
- *or several agencies assigning different language groups in the same country* – such as Flemish, French and German in Belgium.

Two distinct types of registration are happening here:

- registration of *publishers* by the registration agencies;
- registration of *publications* by the publishers!

Because publishers can only realistically be registered on a national level, the ideal is for all agencies to work at the national, rather than the language level. Currently, there are around 155 registration agencies in total.

Top

Linking books with data (and data with other data!)

A small set of *core details* about each book must be collected from the publisher (Registrant)

- as an aid to maintaining uniqueness of the ISBNs,
- and a link between the each ISBN and the publication it refers to:

These are the "minimum referent data" that make a given ISBN unambiguous in pointing to one specific *product*. They are listed in the ISBN manual as:

• ISBN • Product form (e.g. hardback book, paperback, audiobook, ebook…) • Title • Series • Contributor (author, illustrator, translator etc) • Edition • Language(s)	• Imprint • Publisher • Country of publication • Publication date • ISBN of parent publication (e.g. when registering a chapter of a book for separate sale)

These details link the ISBN unambiguously with the product – like the "identifying characteristics" for a criminal suspect as we discussed in "Mistaken Identity".

They also offer a basis for value-added data that either the RA or a bibliographic agency may provide, possibly using the ONIX for Books Product Information format to give a much more detailed, well-rounded view of each publication.

Here is a selection of some known books in print services in EU member states for you to explore:

EU Member state	Books in print service(s)
Belgium	http://www.boekenbank.be/ http://www.banquedulivre.net/
Czech Republic	http://www.sckn.cz/ceskeknihy/
Denmark	http://www.bogguide.dk/
Finland	http://www.kirjavalitys.fi/kv/
France	http://www.electre.com/
Germany (plus Austria, German Swiss)	http://vlb.de/
Greece	http://biblionet.gr/
Italy	http://www.ibuk.it/
Netherlands	http://www.cb-logistics.nl/
Spain	http://www.dilve.es/
Sweden	http://www.bokrondellen.se/
United Kingdom	http://www.nielsenbookdata.co.uk/
Norway	http://www.bokbasen.no

Registration Agencies collect the essential contact details of all their Registrants, together with imprints (brand names) they own.

This allows anyone looking up an ISBN to follow a chain of information back to the original Registrant (if they are still in business). Look-ups may be a paid-for service.

Contributors to books in print services, as the name implies, must provide information about the *availability* of the books in their data. This supply information closes the final link in the

supply chain between the author and reader of the book, with each step enabled or even promoted by the use of ISBN according to the ISO standard.

Linked Heritage partners in Hungary and Germany are ISBN Registration Agencies for those countries, and in Italy, the RA is closely linked to our Linked Heritage partner.

Because each national group has its own agency, and as the Registration Agencies can choose how the provide the ISBN system in each country, there are differences in *how* the numbers are administered, *though not in the service itself* where the standard is followed.

Top

ISBN in Hungary

The Hungarian National ISBN Agency is based in the National Library of Hungary (or NSL, National Széchényi Library, after its founder):

- The Hungarian ISBN and ISMN Office registers publications for both standards
- Registrations are stored in the in-house developed Hungarian ISBN Database
  - In use since 2003, when manual registration became unmanageable due to volume of work
  - The database contains data on printed and electronic materials which are identified by ISBN and published in Hungary.
  - Provides efficient information on documents to be published and lists of ISBNs which help NSL's work on collection of legal deposit copies
- Provision of richer descriptive data on new publications to the ISBN agency will become mandatory under a new legal deposit law
- The agency issues ISBNs to Hungarian publishers free of charge, which implies a cost saving
  - for publishers – it is marginally cheaper to offer books for sale in a supply chain, than in some countries where publishers pay a fee
  - for the agency – there is no overhead to manage ISBN allocation fees

Further developments of NSL's ISBN Database are planned:

- Creating ISBN lists of a publisher's output in various media (printed materials, such as books and maps as well as electronic materials such as compact discs or e-books), for the internal use of the Legal Deposit Department, of which the Hungarian ISBN and ISMN Office is a part
- Deduplication routines to maintain uniqueness according to the ISO standard
- Compatibility with main database of National Széchényi Library, for exchange of bibliographical data
- Replacing manual registration with an electronic form, internally and for publishers:
  - Publishers will be able to assign ISBNs by filling an online form
  - The process will generate a simple bibliographical record at the same time
- External interfaces will include:
  - Providing the Hungarian Statistical Office with statistical information on books
  - Providing data to a future Hungarian books-in-print system
  - Allowing publishers to check the status of deposit copies of their publications (arrived or missed)

Our Linked Heritage partners at NSL identified some misconceptions and mistaken uses of the ISBN which usually come from expecting the standard to deliver more than just an identification system:

- **Misunderstandings** expressed by publishers:
  - Without an ISBN the publication cannot be considered as a book
  - ISBN is connected with copyright and gives protection against illegal duplication
  - In case of introducing a new price, a new cover or a new logo a new ISBN is also necessary
  - Other types of material might be assigned ISBN because ISBN is free of charge in Hungary and it is an easy way to create a GTIN-13 code for another type of product
  - Usage of ISBN is mandatory and publishers must pay a fine if it not used
  - Publishers have no obligations to deliver legal deposit copies of books without ISBN

- If a publication has an ISBN, they have to give an extra copy to the Hungarian ISBN and ISMN Office, in addition to the legal deposit copy
- If a publication has an ISBN, they have to give only one legal deposit copy and they send it to the Hungarian ISBN and ISMN Office.

Failure to apply the ISBN standard itself will result in faulty numbers that do not realise the intended benefits of applying the standard. In NSL's experience:

- **Non-standard (i.e. incorrect) uses** of ISBN by publishers include:
  - Same ISBN for both print and e-book version of a book,
  - Same ISBN for both hardback and paperback version of a book,
  - Same ISBN for two or more different books,
  - Same ISBN for different editions of a book

Top

ISBN in Italy and Germany

The ISBN agencies in Italy and Germany are rather similar so they are described together:

- In Italy and German Registration Agencies are maintained by subsidiaries of the publishers' trade bodies in those countries
- Both RAs also register publications in their respective languages published in other nations (*e.g.* German publications from Belgium, Italian publications from San Marino)
- In both countries, ISBN registration is paid for by the Registrant, usually the publisher (fees are very low, usually less than €1 per ISBN, though they can be higher for small publishers using only a small number of ISBNs)
- Both ISBN agencies offer the **ISBN-A service** detailed in the separate case study

In Germany and Italy, the book industry is highly integrated – there are well-established relationships across the supply chains. This makes application of the identifier standard far simpler – of course, there are difficulties, but it is easier to correct mistakes and introduce new users to the standard.

Top

Lessons from the case studies

- The ISBN illustrates all the features discussed in "Identifiers in the commercial world" and "Mistaken Identity". This is because it is one of the oldest and best established standards, and because it was designed for an industry with a well-developed supply chain. Other identifiers' case studies can be compared with these cases
- An important part of the work of a Registration Agency is educating users of the identifier, and correcting mistakes and misunderstandings – this is even more important in countries or industries where a standard is relatively new. It is not enough just to build a technological platform without support and promotion
- Commercial sector and heritage sector organisations *can* cooperate in providing a successful persistent identifier service – where there are clear benefits for everyone who contributes

Top

EXPLORE FURTHER

From Dickens to Dante: ISBN propels book trade to billions (PDF), by Stella Griffiths; ISO Focus, 2011.
An ISO Focus special report outlining the current state of the ISBN standard, new developments and challenges.

Top

NOTES

1 http://www.isbn-international.org/pages/media/Usermanuals/ISBN%20Manual%202012%20-corr.pdf ↑
2 http://www.abc-clio.com/ODLIS/odlis_m.aspx#monograph ↑
3 http://www.isbn-international.org/faqs ↑
4 http://www.isbn-international.org/faqs ↑
5 Michael P. White © http://www.ala.org/alsc/issuesadv/kidscampaign/clipart ↑

Top

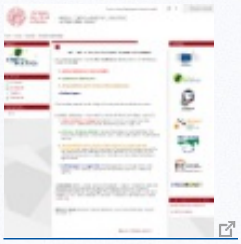
« Back Forward »

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: COMMERCIAL AND HERITAGE VIEWS

4/9 – CASE STUDY 2: THE ISBN-A – MAKING ISBNs PART OF THE WEB

Readings and training materials



Summary

This case study explores how the ISBN system is still useful to the world of E-books and the Web, and how new possibilities are starting to open up for finding and selling books online with ISBN-A.

Page Index

The ISBN system was developed in the 1960s and 1970s to allow reliable, efficient exchange of information for bibliography, stock control, ordering and sales reporting.

This was a triumph of standardisation in the early days of computerised commerce; however, it was never designed for use in a global network of computers like the Internet.

Yet numbering system can still work in the context of the World Wide Web. In fact, recently:

- New ebook guidelines for Registration Agencies made clearer when electronic publications should get new ISBNs, so that both sellers and end-purchasers can distinguish between products that *have a significant difference for someone in the supply chain;*
- Fitting ISBNs into the larger GTIN-13 namespace means that more numbers will be available; these are needed because
  - *More publications will be produced* with digital technology, and
  - *More different versions and formats of each publication will be made*, especially different file formats and “packages” of allowed uses for ebooks.

ISBNs are still mainly *used* through a network of agencies whose organisation and equipment pre-dates the Internet. In many *cases, services to look-up product information and ordering* (as the ISBN’s ISO Standard recommends) have not been fully developed.

ISBN-A, or “Actionable ISBN”, aims to give publishers and others a central, standardised service platform to offer these additional services based on ISBN. It takes advantage of the Internet’s capabilities by integrating ISBN into the DOI system.

Top

Background to the DOI system

To explain what the DOI system is, first a little Internet anatomy:

- The Internet itself is an *infrastructure* of computers connected together to exchange information. It is made up of
  - Computer hardware belonging to public and private organisations, individuals, Governments etc.
  - a standard set of rules (called *protocols*, such as HTTP, FTP…) for how the computers share digital information ("ones and zeroes"). It is neutral in terms of what the information means, and how it is encoded.

One simple analogy could be the system of rails and railway stations connecting cities and countries together. All types of rail transport use it, but it does not tell you what type of trains it carries, when they are scheduled, and what is in the trains (*e.g.* it could be passengers, freight, or both). Connecting the systems depends on *e.g.* the gauge of the rails, signalling rules, avoidance of collisions in scheduling…

- The World Wide Web (WWW) is made up of documents (and other digital files) available through HTTP requests on the Internet, and the links (URLs) between them.

In the railway train analogy, this is the "web" of all the passenger journeys through the railway network you and others could make, and all the meetings you might have at each stop or change – but the passenger analogy is not perfect, because railway trains *transport unique persons*, but the Internet and the WWW only *transmit copies of digital files between computers*.

- Just as there are both passenger trains and freight trains on the rails, the WWW is not the only thing that uses the internet. Other services use the internet at the same time (for example, e-mail services).
- A URL ("Uniform Resource Locator") in the WWW is an *identifier* pointing to the *location (only)* of a digital file (i.e. a piece of digital storage on a computer on the network). It can also contain a command to software running at that location, but we will not consider this complex case.

  This leads to the two most basic problems with retrieving information through the Web:
  - How do I know this specific file (or any file) *claimed* to be at this URL will *actually* be at this location?
  - What information *about* the file (*e.g.* its meaning and intellectual content, its encoding format, history of changes) can I reliably get *before* I download the actual file?

  Clearly the answer to these questions depends on some level of *organised management* of files and information about them (and their contents), and only in the second place, on *technical solution*s for implementing it.

It is important to start with a clear understanding of the limitations of HTTP. There are two

basic types of "things" that can be identified on the Web:

THINGS IDENTIFIED ON THE WEB

Things you can *get* through HTTP	Things you *cannot* get through HTTP
Copies of digital files made available on the Internet. That's all!	Concepts (*e.g.* theories, histories, names, subjects of documents, instructions for use… this includes concepts describing digital files!) Unique physical things (*e.g.* you, me, the Taj Mahal, my laptop, Einstein's brain, the Mona Lisa…) Also: digital files *not* made available on the Internet…
N.B. some of these digital files could be *about* other digital files, or about things that are not digital files – but you don't know that until you *get them and inspect them*.	N.B. most identifiable things in the world are *not* available through HTTP!

In the early days of the Internet and the WWW, there was a plan to use three distinct types of structured identifiers for digital files ("objects") and related information:

- URN – Uniform Resource Name: A persistant, location independent identifier for an object
- URL – Uniform Resource Location: The address of an object, contains enough information to identify a protocol and retrieve the object
- URC – Uniform Resource Characteristics: Any combination of one or more URNs or URLs with meta information [1]

So: a URN is the identifier for *what I am trying to find*, the URL should be the *location I need to get it from*, and URCs make the link between the two, telling me *about* the file and *where* to get it.

The network location (URL) of a digital file (identified by a URN) is itself a "characteristic" (URC) of the file.

In the end, URCs didn't work out, and we now use the term URI (Uniform Resource Identifier) to cover both URNs and URLs. The important point to grasp is that URNs are about the identity of a document or resource, independent of its location, and the URL is about the location of a document or resource (somewhat independent of its identity).

The aim is to link network locations (URLs) of copies of a file to a single, unique name (URN) for that file, and *keeping the location information (which would have been the URC) up to date*.

- The first two, URN for names and URL for file locations, have been linked by national libraries for their digital collections and national bibliographies, as in the URN case study from the Hungarian National Library.
- This "resolution" is like a catalogue, recording a book's ISBN and the fact that a copy is held at a particular library.

More details *about* the file, beyond just "where to find it", are hard to standardise because they need cooperation across a whole sector of activity, and between sectors.

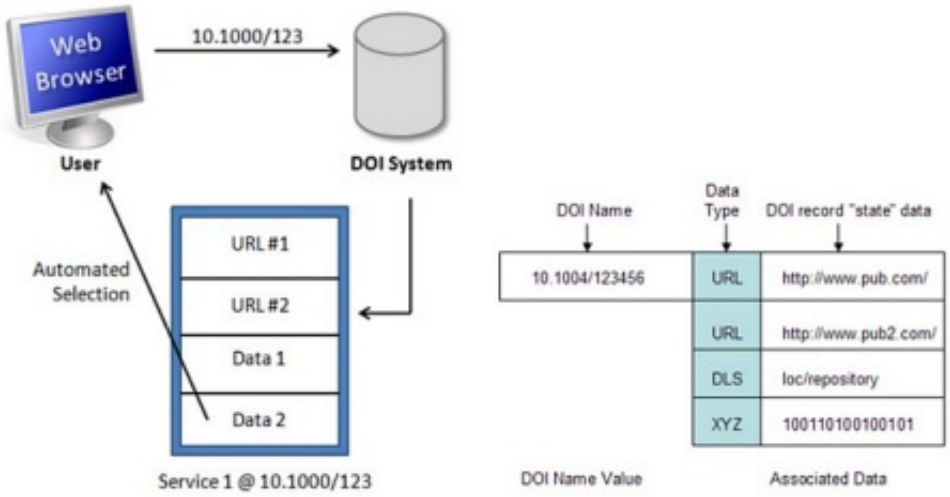
Top

The DOI system

To offer a solution to this problem, the DOI system was launched, offering:

- A *service* for maintaining unique, persistent identifiers, featuring
  - Persistent names – like ISBNs, DOIs cannot be changed or re-assigned once assigned
  - Resolution to one or more URLs, with optional *additional metadata* based on *cross-sector cooperation*

DOIs are unique numbers created and managed by registration agencies in a similar way to the ISBN. Their service platform makes them automatically resolvable on the Web in a much more immediate way:



2

A DOI user sends a *request* to the DOI system at http://dx.doi.org - the DOI system responds with some of the *metadata* stored for that DOI, or with a *copy of the file identified by the DOI*, depending on the type of object the DOI identifies, and the appropriate kind of access available (for example, if a payment is necessary, the user might see a page requesting subscription). A Web browser can automate some of this process.

To keep a DOI resolving to the latest information about the thing it refers to, even when the actual location of the thing is changed, the owner of that DOI has to actively *manage* these links.

- The DOI system provides a *platform* to enable this but does not *automatically* track *e.g.* the current URL of a scientific journal article. If a journal title is sold to different publisher, the new publisher must take ownership of the DOI and the add new URL for the article
- DOI Registration Agencies can specialise in *one type of material*, for a well-defined category of users. This makes it much easier to define "Uniform Resource Characteristics" for those materials

Domain-specific implementations include:

- **ISBN-A** for books (see below)
- **DataCite** for scientific data sets (see case study #3)
- **CrossRef** for academic publications (normally journal articles)
  - Identifies academic publications at the *journal article level*: extremely convenient for precise citations
  - Can identify academic books and chapters of books, and parts of content (*e.g.* tables, diagrams)
  - Citations can be Web-linked through their metadata to become part of the research process
- **EIDR** for film and TV assets
  - A linking ID for existing audiovisual data and identifiers, with links;
  - Identifies films and TV programmes (and series) at any level of abstraction:
    - defines the relevant characteristics of each level
    - different versions (*e.g.* the "director's cut") and encodings (*e.g.* a DVD or digital download) can be linked together

Current DOI implementations are used for published (usually commercial) creative works. But other types of object could be identified and looked up using DOI.

DataCite is a case in point since research data sets do not neatly fall into the category of commercial publication.

- The technical, organisational and financial model of DOI is flexible enough to be used by non-profit organisations:
  - Software platform based on CNRI's open-source Handle system;
  - Some functions of a Registration Agency can be delegated or shared;
  - The business model of a DOI Agency can incorporate for-profit or cost-recovery pricing
- The core data model for DOI metadata is demonstrably compatible with those for heritage data (using *e.g.* CIDOC-CRM).

DOI could be used to identify museum objects (see the "Heritage Sector Web Identifiers" section).

Top

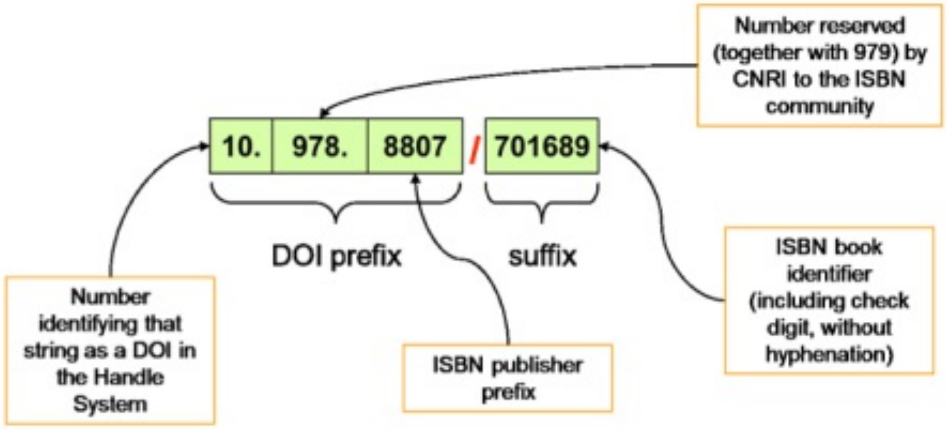
ISBN-A and ISBN

ISBN–A ("actionable ISBN") is an innovative marketing tool to use ISBNs on the Web. ISBN-A initiatives in Europe so far are led by the ISBN agencies of Italy and Germany; the DOI Registration Agency that supports these is Linked Heritage partner, mEDRA (multilingual European DOI Registration Agency).

ISBN-A provides a *platform* for managing and exploiting the metadata for marketing books in the Web environment. Each ISBN Agency provides the service according to its own business model and marketing strategy. Creating ISBN-As for books enables *services* to be built on the metadata and links that publishers provide for them.

To make ISBNs into identifiers in the DOI system, the ISBN is simply incorporated into the DOI syntax.

Let's take an example number: 978-88-07-70168-9. In the DOI system as an ISBN-A it appears as:



3

A Web link is created from the number simply by adding the HTTP address of the central DOI resolver service, to form http://dx.doi.org/10.978.8807/701689 and this link can then be resolved by any Web browser or Web-based application. Of course, what this resolves to - whether to a web page describing the book, or a page where you can buy the book, or the author's home page - depends on how the publisher of the book manages the DOI data.

Top

ISBN-A in Italy

A DOI Registration Agency is domain-based; an ISBN Agency is usually national. So the ISBN-A system (at least in Europe) is based both in Italy (where the DOIs are registered) *plus*

another country if the ISBNs are registered outside Italy.

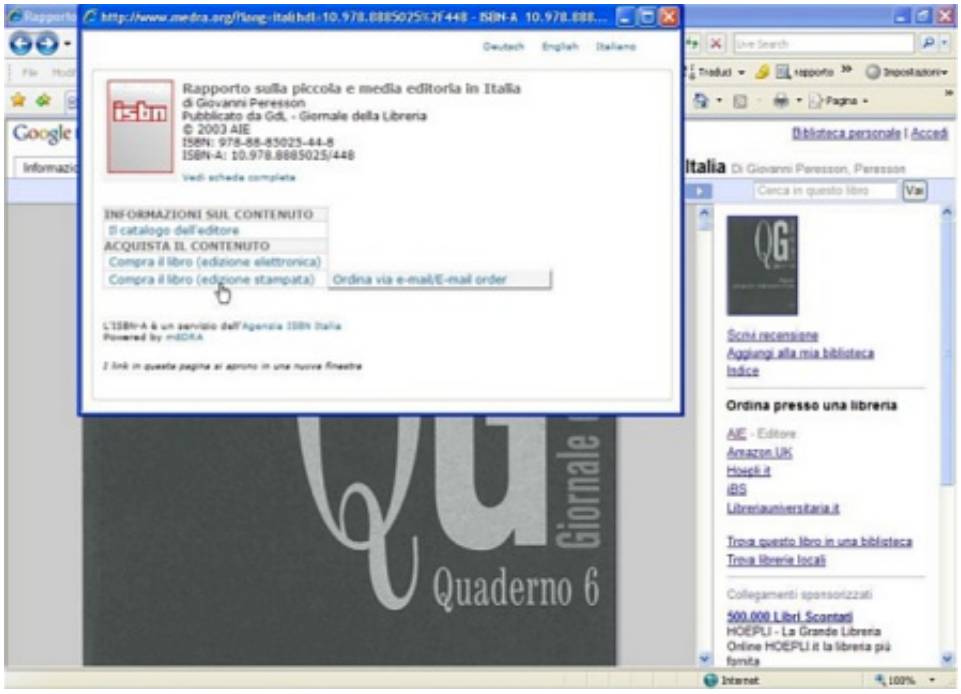
- mEDRA is the technological provider of the Italian ISBN Agency so in Italy the ISBN-A has been fully integrated with ISBN
- ISBN-A registration tools developed by mEDRA for the Agency:
  - Web service to enable publishers to submit ISBN-A registration messages in ONIX format;
  - Online web platform integrated with the Agency bibliographic database to enable publishers to add resolution metadata to ISBN bibliographic records.
- An indirection service is managed by mEDRA, allowing links for multiple retailers to be found on resolving the book's ISBN-A:
  - mEDRA designs the user interface (or "landing page")
  - a "view metadata" service is built on bibliographic data maintained in the local ISBN Agency database
  - book covers can be displayed based on images uploaded to the ISBN Agency platform
- Prices for registration and maintenance depend on how many products are registered: from around €100 for 10, to €1500 for 500.

The ISBN-A platform has potential primarily as a business-to-business (B2B) service. Its value depends on:

- Investment of resources by publishers to maintain and develop the marketing data and links
- Added-value development of the range of options offered by the basic "title page" provided by mEDRA – perhaps partnerships with:
  - social media channels,
  - Web analytics
  - or even content distributors (ebook wholesalers)?

As a single Web link per book with a vast number of optional features possible, the advantages of embedding such identifiers in the context of Web pages, search engine results, social media and mobile applications are clear.

The existing presentation of data to end users is very simple, as in the example below, and primarily text-based:



In principle, the ONIX metadata format supported by mEDRA's resolver could offer links to almost any type of media, depending on the context of resolution, *as long as these were provided and managed by the publishers*.

Top

ISBN-A in Germany

The first full adoption of ISBN-A outside Italy is in Germany. The German publishing association has a services subsidiary, MVB, which runs the German ISBN agency and a variety of online platforms, including a retail Web site and a books-in-print database.

- mEDRA provides the German ISBN Agency with a B2B service for the ISBN-A registration (a Web Service using the ONIX for DOI format)
- The German Agency has developed its own tools to collect ISBN-A registration requests and updates from publishers (*e.g.* via Web form)
- The indirection service and landing Web pages are locally maintained the German Agency.

Because MVB also runs the Web bookshop “Libreka!” it can easily add retail links to ISBN-A book data.

The marketing materials it can assemble as part of running Libreka! and the books in print service “VLB” put MVB in a very strong position to develop bespoke book marketing channels which add value to the data.

Click on this ISBN-A maintained by MVB to see the system in action:

10.978.37639/35147

Top

Lessons from the case study

- The successes of standard identifier systems like ISBN can carry over into the Web era
- All the same costs that come with “traditional” identifiers are implied by managing identifiers on the Web
- But – it is possible to offer cost-effective, persistent identification through central service platforms like DOI
- DOIs can be used to Web-enable existing systems like ISBN
- Web services based on actionable identifiers like DOI really need extra management effort and investment to provide value
- This may carry over to new areas like cultural heritage if there is demand from that sector

Top

EXPLORE FURTHER

DOI Handbook and topical Factsheets
The Handbook gives technical overviews of DOI, and the fact sheets introductory outlines on specific topics

DOI Factsheet on ISBN-A
A summary of the ISBN-A system as a DOI implementation that incorporates an existing standard

Top

NOTES

1 http://tools.ietf.org/html/draft-ietf-uri-urn2urc-00 ↑
2 DOI Handbook ↑
3 mEDRA ↑

Top

« Back Forward »

Linked Heritage LEARNINGObjects

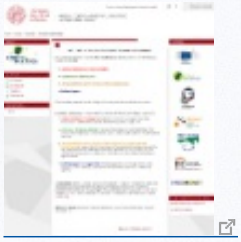
PERSISTENT IDENTIFIERS: COMMERCIAL AND HERITAGE VIEWS

5/9 – CASE STUDY 3: DATACITE AT TIB-UB[1]

Summary

The DataCite service identifies research data outputs so they can be re-used for new research. In this it shares some goals with digital cultural heritage and shows how DOI might be useful to museums and archives.

Readings and training materials



A Linked Heritage partner, the German National Library of Science and Technology (Technische InformationsBibliothek; TIB), funded by the German Federal Government and States, primarily offers a global *documentation service* covering mathematical and engineering sciences:

- Architecture
- Chemistry
- Computer science
- Mathematics
- Physics
- Engineering technology

TIB also hosts a centre of expertise in metadata for multimedia items; exemplified by their PROBADO initiative developing data models for music and digital 3D architectural drawings.

As we discussed in the introduction, identifiers link things with information about their characteristics. Hence TIB, having expertise in research information and the metadata describing it, is a natural home for the DataCite consortium, which maintains DOIs for full *sets of research data*.

Research data was traditionally a kind of “grey literature”, unpublished but still valuable – produced in the course of scientific research.

Though there is not a commercial supply chain for research data (yet?) it is possible to identify a kind of “audit trail” for the products of the academic research process:

[2]

DataCite’s identifier and attendant metadata aim to enable *citation* and *quotation* (or re-use)

of original data in new contexts.

High-powered, networked computing can derive new value from large aggregations of real data through:

- *Simulations* – creation of **model experimental situations** within a computer, based on real-world data
- *Meta-analysis* – **statistical studies** of research data combined from many real-life studies

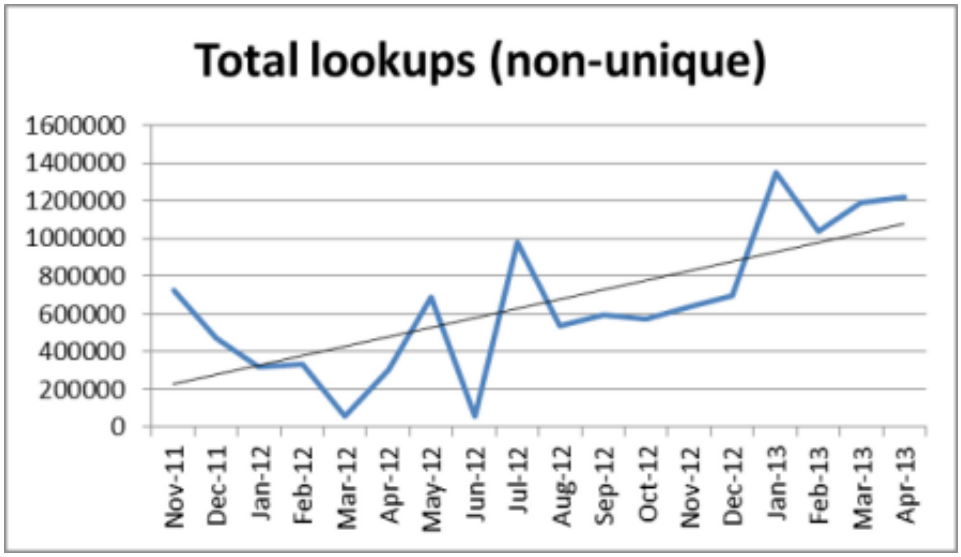
By sharing their original raw data from the earliest stages - rather than only at the end and in condensed and summary form - scientists can:

- do more and better experiments,
- more quickly,
- and get more value from the research that has been funded in the past.

DataCite uses DOI to give the research data identifiers that are:

- **persistent** – the *location* of a data file may change when an academic changes institution, or when data archive systems are changed
- **flexible** – the DOI system can link *many versions* of a data file to a *single identifier* so that researchers can choose the most appropriate copy, *e.g.* by format, subset of data or language of text

The DataCite service reports [3] that it has registered 1,498,811 DOIs to date; there has been a steady growth in (successful) look-ups since the end of 2011:



The requirements for **persistence** and **flexibility** are identical to those in the digital cultural

heritage world:

- a digitised version of a cultural heritage object, or documentation about it, may move in the same way as a research data set
- researchers and interested members of the public may want to access different resolution images, or description in different languages

Top

Heritage datasets in DataCite?

There are some real similarities between the data available through DataCite and those aggregated and curated by cultural heritage projects like Linked Heritage:

- Both begin with some kind of *unique event* which is the object of interest, rather than a set of information – or tangible object that witnesses to an event or a series of events in history
- Both types of data *arise from "research"* – in fact, a set of related Linked Heritage Cultural Heritage Object records could be considered a dataset in DataCite terms
- Both types of data are intended for *use in further "research"*
- Additional value can be added to both types of metadata –
  - through linking multiple digital surrogates,
  - and linked the metadata to object data through RDF representation

Several heritage institutions have already used DataCite to assign DOIs to some of their digital collections; there are almost 19,000 documents submitted to the DataCite metadata catalogue whose publisher is some kind of museum. Most of these come from the Museum of Vertebrate Zoology at Berkeley as might be expected (they are in fact specimen records from a natural sciences collection).

DataCite also contains a set of web pages for paintings by Karl Hagemeister at the Bröhan Museum, Berlin (one example: Birken am Bach im Spätherbst). Possibly collections like these are under-represented in DataCite due to

- existing methods of citing heritage objects like paintings
- slight differences in emphasis in the type of "research" that uses museum data:

	DATACITE RECORDS	LINKED HERITAGE RECORDS
Object of description	Unique "event" (experiment, investigation, publication or review)	Unique curated cultural heritage object or collection
Data creation context	Data created in the course of scientific research on a specific problem or project theme	Data created in the course of on-going curatorial research into gallery, museum, archive or library's cultural programme, or one specific project theme
Access to source object	Full original data often available openly	Original heritage object normally accessible to the public (or at least bona fide researchers)
Richness of metadata	Basic metadata to enable unambiguous identification, relationships with other identified objects, interoperability with richer data schemas	Extremely rich metadata enabling (in principle) detailed description of described object *AND* all related entities
Access to metadata	Metadata openly available (or	Metadata in full LIDO format

describing object	notice as to why not)	*NOT* yet openly available
Digital surrogates available	Multiple representations of data and/or metadata possible through content negotiation	Multiple Digital Objects (*e.g.* digital photos of a building taken from different angles, scans of different pages – or different sections of pages – of documents) often collected and linked via on LIDO record
End-user value of metadata and source object	Citation and re-use (quotation, further analysis, incorporation into meta-study) of datasets in academic research, educational materials	Citation of object records in academic and professional research, educational materials, public-service cultural offerings
Underlying data model	Indecs-compatible data model (DOI Kernel)	CIDOC-CRM compatible data schema
Potential for linked data	Linked data representation of DataCite metadata already published	LIDO as RDF in development

Top

Lessons from the case study

- The DOI system is already used for a type of data that is very similar to heritage object data
- There are some differences between research data and museum and archive documentation that mean a different solution for this data might be needed
- Both sectors share a concern for digital preservation and this is one reason to use DOI

Top

EXPLORE FURTHER

DataCite: *Information for potential clients (UK via the British Library)*
Step-by-step videos showing how the DataCite system works from the user's perspective

DataCite workshops at the British Library
Presentations providing background and context, including case studies of DataCite implementations such as the Archaeology Data Service, UK, Data Archive, and Data.bris in the UK, as well as discussion of technical issues and possible future developments

Top

NOTES

1 This section draws on Brase, J. (2012) *DataCite and Linked Data*. Presented at: Global Interoperability and Linked Data in Libraries, Florence, Italy, 18th-19th June 2012. ↑
2 *Data, information and knowledge in the academic research "trajectory"* (after Brase, J. 2012). ↑
3 See http://stats.datacite.org/ for full statistics ↑

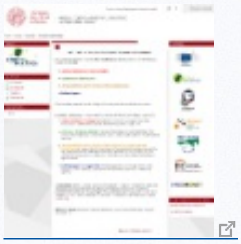
Top

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: COMMERCIAL AND HERITAGE VIEWS

6/9 – CASE STUDY 4 – URNs AND URIs

Readings and training materials



The URN mentioned in the DOI case study is a type of identifier used by the NSL (National Széchényi Library), a Linked Heritage partner, in addition to ISBN.

- ISBN is itself a registered URN namespace
- ISBN-A works through the centralised DOI resolution service, which is the same as a URN resolver
- But – there is no single unique ISBN URN resolver

As we discussed in the DOI case study, a URN is used to "name" an object of interest:

- In principle, it can be used – as the ISBN frequently is – to link to information about the resource
- In practice, URN resolution services are most often used simply to provide *persistence of URL linking* – that is, to avoid broken Web links
- Since libraries like NSL are the most prolific users of URN resolvers, there is usually *some descriptive information* at the URL provided by resolving the URN
- This may not (necessarily) be in any kind of "linked data" form.

URN as a guarantee of URI persistence

URN resolvers such as the one at NSL essentially provide one item of metadata for each URN: an up-to-date Web location, or URL. This is - in effect - the same sort of resolution service as provided by the very simplest application of a DOI, where one unique ID resolves to one URL. And the resolver requires the same type of management when the location of the object is moved for any reason.

URN resolution at NSL

The URNs created by NSL are integrated as a persistent ID in the NSL's online catalogue, making the metadata links independent of changes to their internal file servers.

There are two main types of materials in scope for these identifiers at NSL:

- Ebooks in the Hungarian Digital Library
  - 11,000 ebooks, including audiobooks, digitised and born-digital books are identified by URNs
- Digital images in the Hungarian Digital Image Library
  - 70,000 images are identified by URNs

No materials from outside the NSL are given URNs.

EXPLORE FURTHER

DOI® System and Internet Identifier Specifications
Compares URN, URL and DOI as identifiers and explains some of the issues raised in this case study

Lessons from the case study

- URN resolvers are useful for identifying local resources and making their links permanent
- URN does not offer much advantage over DOI

« Back Forward »

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS: COMMERCIAL AND HERITAGE VIEWS

8/9 – CONCLUSIONS

Thank you for viewing our Learning Object *Persistent Identifiers: Commercial and heritage views*.

We invite you now to answer some questions for your learning assessment and to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: Persistent Identifiers: Commercial and heritage views.

For further details:

Readings and training materials

« Back Forward »

Linked Heritage LEARNINGObjects

PERSISTENT IDENTIFIERS:
COMMERCIAL AND HERITAGE VIEWS

9/9 – CREDITS

Project by:
Michael Hopwood
Graham Bell
Paola Mazzucchi
Giulia Marangoni
Petra Mensing
Istvan Moldovan
Szabolcs Dancs
Stella Griffiths
Michael Vogelbacher
Natascha Pattock
Patrizia Martini

Graphic and Web design by Gianluca Drago
CAB, University of Padua (Italy)

Funded by the Competitiveness and Innovation Framework Programme, 2011.

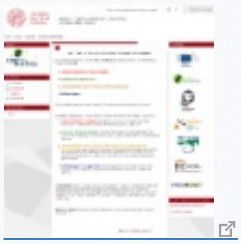
« Back

Linked Heritage LEARNINGObjects

PUBLIC-PRIVATE PARTNERSHIP WITH EUROPEANA

1/3 – PUBLIC-PRIVATE PARTNERSHIP WITH EUROPEANA

Readings and training materials



You are interested in the possibility of partnership with Europeana, and wish to find out more about:

- What Europeana is and does;
- Why it may be of value to the commercial sector;
- Any relevant legal questions;
- Technical services available to support collaboration.

This brief overview gives answers those questions, with links to more detailed information, drawing on the work of EDItEUR in the Linked Heritage project, 2011-2013.

- Anyone considering commercial partnerships with the cultural sector more generally may also find useful ideas here.

Page Index

Europeana and the Europeana Network

Europeana is the European Commission's flagship **digital cultural heritage portal**. It can be found at:

www.Europeana.eu

It is a central discovery and access point for European historical, social, cultural and artistic materials that have been made available online in digital form by (mainly) public institutions, such as

- Museums
- Art galleries
- Libraries
- State archives
- Audio-visual (film) archives
- Sound archives (music & other recorded sound)
- Photo libraries and agencies

These organisations host physical cultural objects, and may digitise them to make them more widely accessible. They also host newer, born-digital cultural items. They then contribute **metadata** describing their digital collection (**not** the materials themselves) to Europeana, along with small **thumbnail preview images**. Some other types of preview can be supported, such as sound clips.

Contributors can optionally join the voluntary **Europeana Network** of metadata providers and heritage / information technology experts.

Freely accessible to the public, and entirely **publically-funded**, Europeana itself primarily aims to offer

- **Aggregation** of metadata about digitised cultural content (again, **not** the content itself)
- **Facilitation** of mass digitisation and expertise-sharing projects in the heritage sector
- **Distribution** of the metadata about existing content (but **only** the metadata)
- **Engagement** with European digital culture in new ways developed by Europeana Foundation's projects and those of its network of contributors

More technical details about Europeana can be found at the Europeana Professional sub-site.

Some more commercial organisations have to date taken part in **mass digitisation programmes**, in partnership with public institutions.

This usually means providing:

- initial investment and perhaps
- technical services

...in return for:

- access to the cultural items themselves,
- temporary exclusivity over sales or licensing of the digital content produced,
- or (more rarely) a quid-pro-quo of content enrichment.

Most commercial organisations that have digitised materials and shared metadata with Europeana to date have:

- digitised **public domain (out-of-copyright) content**,
- or shared **metadata which is largely descriptive** and typically **lacks** either
  - Richer previews of the content (*e.g.* cover images or extracts), or
  - Links to retail pages to buy some product, a commercial version of the digitised content.

A larger scale collaboration would aim to include

- both out-of-copyright material and in-copyright culturally-relevant products, with
- **both** richer preview content,
- **and** links to (at least) one retail landing page per product

This would realise both

- fuller cultural enrichment of the public database (for Europeana and its users),
- and commercial sales potential (for the metadata provider).

and would form the basis of a mutually-beneficial partnership.

Top

Benefits and costs of contributing to Europeana

Three fundamental conditions for a commercial partnership with Europeana are:

1. *Commercial incentives* (financial or otherwise) for the data provider or their clients;
2. *Technical availability* of data to keep costs to a minimum;
3. *Ownership of rights to redistribute* data and preview images to Europeana without restriction of any type of reuse of the data.

Commercial incentives are most likely to exist where:

- *One company owns all relevant rights* in products, metadata and preview images, and most importantly, controls all *retail offerings (sales rights)* for all products;
- Organisations have a *public mandate to promote heritage (e.g.* a national library that also publishes new books);
- There are data intermediaries able to reduce technical barriers, and costs, and which can aggregate large numbers of records from their clients (often smaller or heritage sector organisations);
- *Data aggregators* might contribute larger numbers of records if *product identifiers can be resolved* to link an arbitrary number of retail offers ensuring fair competition.

These incentives are listed in order of *increasing value to Europeana* and online heritage.

They also represent *increasing levels of public or private investment* in management of data and development and maintenance of infrastructure to realise that value.

Any sustainable commercial contribution (*i.e.* beyond a one-off prototypes as achieved by Linked Heritage), will require Europeana to invest in enhancing its own commercial attractiveness:

- Appropriate *licenses* for commercial data, to avoid conflicts with existing commercial data supply relationships
- *Data management support* following *commercial best practice;*
- Raising the *commercial profile* of the Europeana portal and data feeds *e.g.* through
  - enhancement of the core data model,
  - alternative API and linked data re-use licences, and
  - support for licensing heritage *content* such as images, full text, sound and video (*i.e.* leveraging Europeana's central position in the cultural heritage sector to simplify commercial access to digital cultural material held by the originating institutions). Easier licensing of digital cultural material is one of the key elements of value that Europeana could offer to commercial organisations, but at present Europeana offers only metadata describing that digital material.

Financial incentives and support may come from public funds or not-for-profits who need product data for:

- Promotion of national heritage and culture,
- International dissemination of literature and other arts in translation,
- Enriching data with industry or trade body awards and events,
- Promoting legitimate retail offerings (anti-piracy),
- Extension of standards and identifier adoption across commerce and archival materials;
- New databases of currently in-commerce products to benefit trade and culture.

Top

Legal context of contributing metadata

One of Europeana's key roles is the collection and further redistribution of metadata. Contributors to Europeana, especially commercial organisations, should attend to the rights status of their that data, in three areas:

Copyright

- Marketing and supply chain data may often contain some texts protected by copyright, for example previews of the content of a product. These *must* be dilligently rights cleared, or removed.
- A commonly expressed view is that individual "facts" in a dataset are not covered by copyright:
  - This *may* help to define *a minimum set of information that may be offered publicly*, and perhaps redistributed as structured data for automated reuse, but -
  - A minimum set will be less commercially relevant and of little cultural depth.

Database right

- In Europe a *sui generis* database right applies to all **collections** of data;
- This applies no matter if copyright protection applies to each **item** of data;
- Hence the right to redistribute and make public without restriction must be checked.

Existing and future licenses (contracts)

- Commercial data intermediaries possess significant assets in the above categories of intellectual property.
- As well as being legally protected, they generate a return on the investment required to create and maintain those assets, through licensing their data and offering related services.
  - New data agreements - such as contributing data to Europeana - must be considered from a business development point of view for the extent to which they offer a return (financial or otherwise) on the investment required to maintain them.

These are all affected by the Europeana "Data Exchange Agreement" (DEA), which stipulates:

> - "All [textual] metadata provided to Europeana will be published by Europeana as open data under the terms of the Creative Commons Zero Public Domain Dedication (CC0)."[1]
> - "This [the CC0 redistribution] does not apply to the content (including previews)… The rights to the content made available via Europeana remain under the control of the data providers who make these objects available."[2]

That is, any contributions of *metadata* from a commercial organisation to Europeana currently involve waiving all rights to the data (both copyright and database right). Of course, this does not affect the rights to the actual content of the products described by that metadata.

- See the workflow diagrams in the next section ("Technical support…") below for clarification of the distinction between "textual metadata" and "content".
- The full terms of the CC0 legal instrument can be found at: http://creativecommons.org/publicdomain/zero/1.0/legalcode
- More explanations are found at http://creativecommons.org/about/cc0 and http://wiki.creativecommons.org/CC0_FAQ

Issues from CC0 for commercial uses

For commercial data owners CC0 raises three main issues:

- In many cases, this will mean that only a slim subset of the data, with an extremely limited:
  - *quantity or selection* of records, and
  - *depth or richness* of information

  can be transferred into the public domain, without undermining the commercial value of the data that made it possible to create and manage the data in the first place.
  - Of course, the lack of richness of the data also undermines the likely commercial value of any reuse.

- Data contributors who do not own the appropriate rights in the data (perhaps because it contains third-party material), must in good faith clear these rights before applying CC0 – this could be expensive and time-intensive.
- Even when all appropriate rights *are* cleared before application of the CC0 licence, the CC0 legal code
  - explicitly disclaims all guarantees of prior third-party rights clearance, and
  - offers *no legal protection or guarantee to potential reusers of the data* that third-party rights have been cleared.

  Commercial re-users of data redistributed by Europeana would clearly desire such provisions, as without them a diligent check for third-party inclusions is required.
- Data licensed under CC0 offers no potential for even limited exclusivity to potential reusers, which might encourage more ambitious commercial exploitation of the material.

These problems could potentially be overcome with alternative licensing arrangements for commercial data contribution, and for redistribution and re-use of the Europeana data that includes commercially-contributed data.

Top

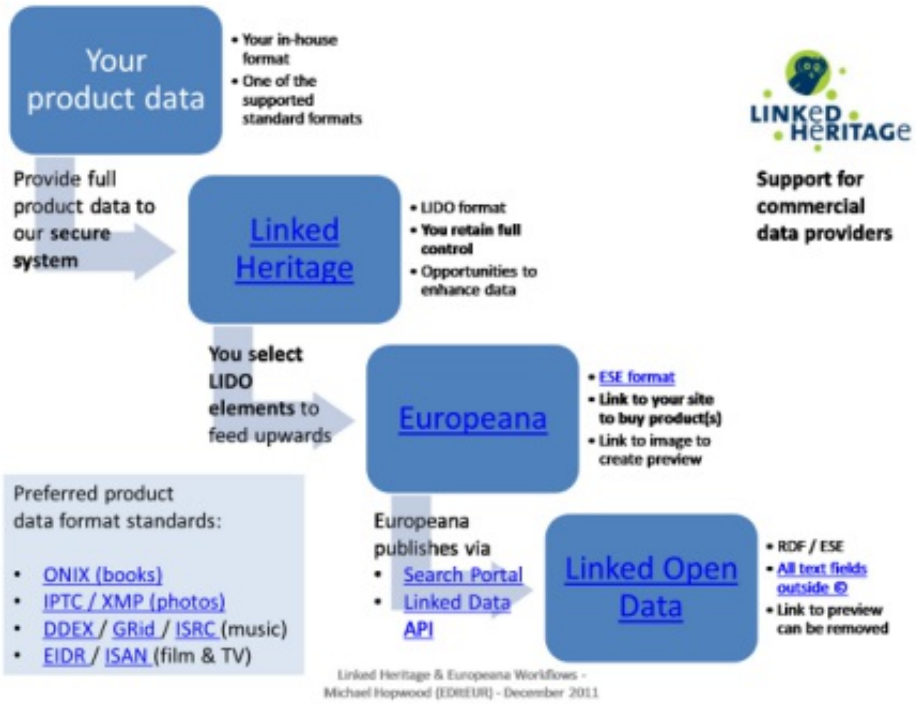
Technical support for commercial partners

The Linked Heritage project ended in 2013, and the products of its research on public-private partnership are *freely available.*

They form a *basis of technical expertise on which to build future data contribution collaborations* between commercial organisations and Europeana and other heritage bodies:

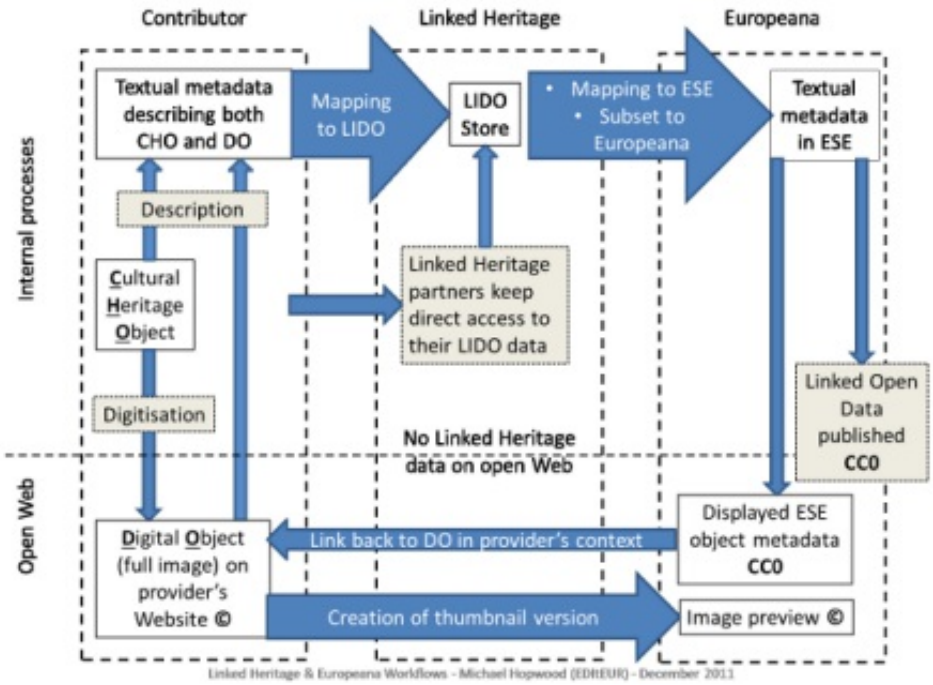
- The first diagram below summarises the *key benefits* offered by this approach, listing industry standards supported and core activities at each stage of data aggregation;
- The second diagram shows the same workflows again from the *technical perspective* of data flows and access to content.

The Linked Heritage *software* and *standards* infrastructure is *freely available* for commercial or public sector use and *further development* and *customisation*.



EDItEUR and its Linked Heritage Work Package 4 partners demonstrated an effective, standards-based approach to this problem:

- **D4.1 Best Practice Report** (link below) introduces *product data standards and best practice of the media industries* for those heritage organisations learning about parallels in the commercial world, and for commercial organisations thinking of *adopting these standards* for the first time;
- *Semantic mappings of industry data formats to the LIDO heritage format* in **D4.2** below were produced in agreement with the relevant standards bodies and hence constitute "standard mappings" that can be used independently of the Linked Heritage project;
- *The LIDO format* is a widely-used "hub" format for data aggregation, and provides a stable mid-point for data collection and enrichment before transforming to the Europeana schema (which has in the past and may in future change significantly), or to other schemas.
- XML *data ingestion and mapping* to LIDO is supported by the open-source MINT software.



Top

EXPLORE FURTHER – TECHNICAL SUPPORT

Linked Heritage D4.1 – Public-Private Partnership Best Practice Report (PDF)
The first part of EDItEUR's Linked Heritage work, describing the major identifier and metadata standards

Linked Heritage D4.2 – Specification of the technologies chosen (PDF)
The second part of EDItEUR's Linked Heritage work, specifying the mechanics of how typical commercial-sector metadata can be mapped into the LIDO metadata scheme used by Linked Heritage

Commercial and cultural sectors: potential for data collaboration? (PDF)
A conference paper presented by Graham Bell, describing similarities and differences between library and commercial book data models, existing commercial sector partnerships with the library world, and potential for future collaboration, especially in the field of Linked (Open?) Data and the Semantic Web

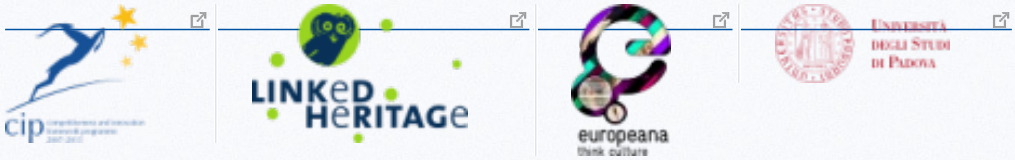
Top

NOTES

1 http://pro.europeana.eu/web/guest/licensing ↑
2 http://pro.europeana.eu/content-and-metadata ↑

Top

Forward »



Linked Heritage LEARNINGObjects

PUBLIC-PRIVATE PARTNERSHIP WITH EUROPEANA

2/3 – CONCLUSIONS

Thank you for viewing our Learning Object *Public-private partnership with Europeana*.

We invite you now to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: Public-private partnership with Europeana.

For further details:

Readings and training materials

« Back Forward »

Linked Heritage LEARNINGObjects

PUBLIC-PRIVATE PARTNERSHIP WITH EUROPEANA

3/3 – CREDITS

Project by:
Michael Hopwood
Graham Bell
Paola Mazzucchi
Giulia Marangoni
Petra Mensing
Istvan Moldovan
Szabolcs Dancs
Stella Griffiths
Michael Vogelbacher
Natascha Pattock
Patrizia Martini

Graphic and Web design by Gianluca Drago
CAB, University of Padua (Italy)

Funded by the Competitiveness and Innovation Framework Programme, 2011.

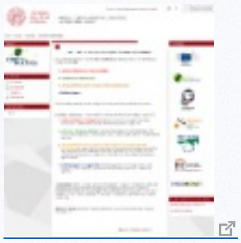
« Back

Linked Heritage LEARNINGObjects

TERMINOLOGY

ABOUT THE LEARNING OBJECT

Readings and training materials



Summary

This Learning Object has been developed in the framework of WP7 of the Linked Heritage project.
Its aim is to introduce work package 3 of the Linked Heritage-project to unskilled or entry level learners in galleries, libraries, archives and museums.
Work package 3 is dedicated to the exploitation of monolingual or multilingual vocabularies in a digital environment. Almost every cultural institution nowadays uses one or more in-house vocabularies for various museological functions. Thesauri are used as an indexing tool, e.g. the subject index at the end of a book; as an inventory device, allowing structural and logical classification of objects in a collection management system. They can be activated in a database or in an online catalogue as a search tool, allowing more detailed and accurate query results. Vocabularies can be implemented as an addition browsing or navigation tool or as an interactive term suggestion application. They also facilitate the synergy of multiple databases.

As the world wide Web is gradually growing with information, specialists have emphasized the importance of vocabularies as a way to structure information on the Web. The knowledge that is inherent to structured vocabularies can help in information retrieval and online searching, thus increase the development of the Semantic Web or Linked Open Data.

This tutorial will explain in a simple and practical way following subjects:

- What is a terminology?
- How a cultural institution can benefit from making their terminologies available in a semantic way?
- What is SKOS?
- What can you do with the Terminology Management Platform?
- More information

Content sections:

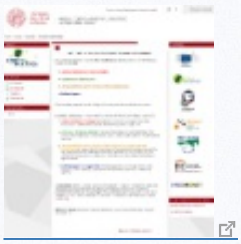
Forward »

Linked Heritage LEARNINGObjects

TERMINOLOGY

1/4 – TERMINOLOGY AND TERMINOLOGY MANAGEMENT

Readings and training materials



We use in this context the word “terminology” in order to mention the resources used by the museums for describing their collections, this word “terminology” might be ambiguous. Indeed, strictly speaking, “terminology” is a discipline which aims at studying terms and their use within a specific domain; but a “terminology” could refer to the resource resulting from this discipline as well. However, “terminology” is the most generic and clear word to mention the different existing types of resources.

Content sections:

Terminology?

The type of resource is highly connected to its purpose: in other words, an information retrieval tool and a knowledge management tool won’t use the same kind of resources. The terminology resources in the context of cultural institutions are mainly used for indexing and information retreival pupose.

Considering this, we have raised five main types of resources organised according their level of complexity:

Simple list of terms

The simple list of terms could be assimilated to a controlled vocabulary. A **controlled vocabulary** is a list of terms that have been explicitly enumerated. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary should have an unambiguous, non-redundant definition. However the simple list of terms generally consists in an alphabetical list of terms of a specific domain without definition or relations between terms… It could be also a list of named entities such as authors’ or persons’ names, location names… It represents the “minimalist” type of resource.

Glossary

A glossary is an alphabetical list of terms of a specific domain where each term has a definition or an explanation. The glossary, despite some common features, is not a dictionary or a lexicon. It often concerns a very specific or technical domain and is generally dedicated to non-experts for giving definition of very technical terms in a simplified way. A glossary could be multilingual.

Classification

Classifications are originally specific to library science and mainly used for cataloguing: a classification is a system of coding and organizing the knowledge. Classification is one of the tools used to facilitate subject access to collections. Thesauri and subject heading systems are another tool facilitating subject access. The main difference between these two tools is that classifications don’t allow assigning an object to several classes while

thesauri allow assigning several terms to one object.

The Dewey Decimal Classification (DDC) and the Universal Decimal Classification (UDC) are the most known classification systems in the Information science and documentation world. DDC is more likely to be used as a system of location of resources while UDC which is more expressive than DDC especially with the relations between subjects will be preferred for subject browsing. Classification schemes may be either special, e.g. limited to a specific subject; or general, e.g. aiming to cover all subjects equally ('the universe of information').

Taxonomy

The taxonomy is very close to the classification since it is also a system of coding and classification. Originally used to designate classifications in the natural sciences field, the word “taxonomy” now refers to a form of classification scheme. In other words, taxonomy could be assimilated to a controlled vocabulary organized into a hierarchical structure. The terms are connected through a parent-child relationship. As classification and taxonomy are very similar, these two types of resources have been brought together for the needs of this report.

Thesaurus

A thesaurus could be defined as “a networked collection of controlled vocabulary terms”. Thesauri allow connecting the terms via several types of relationships which can be hierarchical, associative, equivalence or definition. This means that a thesaurus uses associative relationships in addition to parent-child relationships. A parent-child relationship is expressed by a Broader Term (BT) / Narrower Term (NT) feature. Associative relationships in a thesaurus such as “Related Term” (RT) (e.g. term A is related to term B) are used to express relationships that are neither hierarchical nor equivalent. Equivalence is expressed by the USE (e.g. preferred term) / Used For (UF) (e.g. non-preferred term). Additional information such as definition or remark can be included in a Scope Note (SN). The equivalence relationship is especially useful within multilingual thesauri. Thesauri contain two different types of terms: descriptors and non-descriptors. The descriptors are the terms used for indexing. The non-descriptors refer to all the terms connected to the descriptors through the relationships mentioned above. Non-descriptors are not used for indexing.

A thesaurus can be either monohierarchical or polyhierarchical: in a monohierarchical thesaurus, a descriptor can be connected to a broader descriptor whereas several broader descriptors can be parent of a descriptor in a polyhierarchical thesaurus. This horizontal level of relationship makes the main difference between thesaurus and taxonomy.

A specific norm, ISO 25964-1 Thesauri for Information retrieval has been established in 2012 in order to tackle the evolution of the thesauris in relation with the one of semantic technologies, namely the SKOS format.

Ontology

An ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts. Ontologies are the main kind of resource used for the Semantic Web or Knowledge management as a knowledge representation. The concepts are linked together by hierarchical relationships in one hand and semantic relationships in another hand.

Top

Semantic Web, Linked Data and You

The Semantic Web (part of Web 3.0) is “the Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it”1. It provides “a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by World Wide Web Consortium (W3C) with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. It was proposed by World Wide Web inventor Tim Berners-Lee”.

As we can read on Wikipedia:

> *“Semantic Web is a term coined by World Wide Web Consortium (W3C) director Tim Berners-Lee. It describes methods and technologies to allow machines to understand the meaning - or "semantics" - of information on the World Wide Web.”*

The availability of machine-readable metadata would enable automated agents and other software to access the Web more intelligently. The agents would be able to perform tasks automatically and locate related information on behalf of the user. While the term “Semantic Web” is not formally defined it is mainly used to describe the model and technologies proposed by the W3C. These technologies include the Resource Description Framework (RDF), a variety of data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples), and notations such as RDF Schema (RDFS) and the Web Ontology Language (OWL), all of which are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain.

The Semantic Web is then an evolution of the Web that implies that we, as users of the Web, change the way we make our data/documents available in order to ensure a machine-readable access and human as well.

The Linked Data is a practice of the Semantic Web since once the data is made available online the links between these data is necesary to make sense.

As a first definition we can say:

> *“In Semantic Web terminology, Linked Data is the term used to describe a method of exposing and connecting data on the Web from different sources. Currently, the Web uses hypertext links that allow people to move from one document to another. The idea behind Linked Data is that hyperdata links will let people or machines find related data on the Web that was not previously linked. The main point is that the focus is more about data and how to create and maintain links between these data than documents and links between documents.”*

Here is a more “official” definition from Tim Berners-Lee:

> *“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.*
> *Like the web of hypertext, the Web of data is constructed with documents on the Web. However, unlike the Web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF,. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the Web grow:*
>
> 1. *Use URIs to identify things (anything, concrete or abstract things, not just documents)*
> 2. *Use HTTP URIs so that people can look up those things.*
> 3. *Provide useful information using standards (RDF*, SPARQL) when someone looks up a URI*
> 4. *Include links to other URIs (RDF links generally) to enable the discovery of related information.”*

Digitisation is a long and expensive process which final aim is to make available onling the digital cultural heritage of your institution. Using the Semantic Web & Linked Data technologies to put your data online is a guarantee to exploit the best your digitised content and optimise their visibility o the Web.
Thesaurus and other terminology resources are mainly used for indexing and organising the collections of an institution. The Semantic Web technologies allow linking several different thesauri and institutions and users will have then the possibility to expand search functionalities through federated searching of multiple controlled vocabularies and linked

data sources.

The semantic enrichment provided by the cultural institutions will facilitate multilingual information access and retrieval thanks to a semantically rich visualisation of thesauri and links in between.

How to join the semantic web: guidelines

The WP3 of Linked Heritage has produced a booklet with a set of recommendations and guidelines for bringing your terminology resources towards the Semantic Web.

This chapter is a step by step guide to publish your terminology as part of the Semantic Web.

Content sections:

STEP 1: Conceive your terminology

Building your terminology is the foundation for all the rest. It determines the operations you shall do later when you will make your terminology interoperable with other resources, and when you will link it to a network of terminologies.

The terminology type that we consider “ideal” is a domain-specific, multilingual and user-oriented thesaurus. The closer to the ideal form your terminology is, the more optimised the exploitation of your semantic descriptions on Europeana will be.

Define your collection domain(s) http://www.athenaeurope.org/athenawiki/index.php/A1

Identify your users’ expectations (about your semantic descriptions) http://www.athenaeurope.org/athenawiki/index.php/A2

Define your connection with the datamodel http://www.athenaeurope.org/athenawiki/index.php/A3

Choose the terms for the semantic description of your digital resources http://www.athenaeurope.org/athenawiki/index.php/A4

Organise your terms into a thesaurus structure http://www.athenaeurope.org/athenawiki/index.php/A5

Find equivalend terms in other languages
http://www.athenaeurope.org/athenawiki/index.php/A6

Implement your thesaurus http://www.athenaeurope.org/athenawiki/index.php/A7

STEP 2: Make it interoperable

Evaluate how far skos is compliant with your terminology features
http://www.athenaeurope.org/athenawiki/index.php/B1

Roughly skosify your terminology http://www.athenaeurope.org/athenawiki/index.php/B2

Define with precision the labels expressing concepts
http://www.athenaeurope.org/athenawiki/index.php/B3

Identify your concepts and validate the structure
http://www.athenaeurope.org/athenawiki/index.php/B4

Ensure the documentation of concepts
http://www.athenaeurope.org/athenawiki/index.php/B5

Map your concepts http://www.athenaeurope.org/athenawiki/index.php/B6

Map your (multilingual) terms http://www.athenaeurope.org/athenawiki/index.php/B7

Validate your skosification http://www.athenaeurope.org/athenawiki/index.php/B8

STEP 3: Link to a network

Define the metadata of your terminology
http://www.athenaeurope.org/athenawiki/index.php/C1

Identification of resources for mapping
http://www.athenaeurope.org/athenawiki/index.php/C2

Mapping with other resources http://www.athenaeurope.org/athenawiki/index.php/C3

Validation of the interoperability http://www.athenaeurope.org/athenawiki/index.php/C4

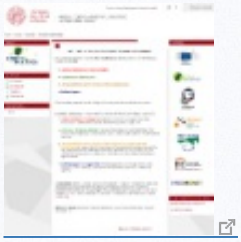
« Back Forward »

Linked Heritage LEARNINGObjects

TERMINOLOGY

2/4 – INTRODUCTION TO SKOS

Readings and training materials



As a Semantic Web compliant format, SKOS is concept-oriented. This means that the fundamental element of a terminology designed in SKOS is the concept and not the term that expresses this concept. The SKOS data model consists of a basic structure that can be extended by specific classes for detailing lexical parts or semantic relations between the concepts of the terminology. The SKOS reference publication summarizes the main features of the SKOS model as follows:

> *"Using SKOS, can be identified using URIs, with lexical strings in one or more natural languages, assigned (lexical codes), with various types of note, and organized into informal hierarchies and association networks, aggregated into, grouped into, labeled and/or ordered, and to concepts in other schemes."*

SKOS data are expressed as RDF triples. This means that concepts may be subject or object and related via a SKOS property which would be the predicate. As RDF triples, SKOS concepts van be identified using URIs. These URIs can be defined according standard persistent identifier systems. The SKOS data model doesn't require the use of persistent identifiers but in a Linked Open Data perspective, their use is highly recommended. Persistent identifiers will be described more precisely in the following sections.

The SKOS datamodel consists in three main components: classes, properties and relations. These three components always start with the prefix "skos:". The distinction between a class and a property is done through the case: the element following the "skos:" prefix starts with an upper-case character when it is a class, e.g. skos:Concept and skos:ConceptScheme are classes; if the element following the "skos:" prefix starts with a lower case character, this means that the element is a property and not a class. For example skos:prefLabel is a property.

Content sections:

SKOS main features

SKOS: concept

SKOS is a concept-oriented model therefore the concept is the central element of the terminology. From a terminology point of view a concept can be defined as an idea, notion or unit of thought. A concept in SKOS is introduced as a class skos:Concept.

SKOS concepts can be brought together into two classes:

- SKOS concept scheme
- SKOS collections

SKOS: concept scheme

A concept scheme is a way to bring together several concepts. A concept scheme is introduced by the class skos:ConceptScheme. An individual concept scheme roughly corresponds to the notion of an individual thesaurus, classification scheme or any other knowledge organization system. It is important to mention that a same concept can be part of more than one concept scheme.

SKOS: collections

A collection is a group of SKOS concepts. A collection is introduced by the main class skos:Collection. Although another class skos:OrderedCollection can also be used in the case where the order of the concepts within the collection has an importance. The notion of collection is different from the concept scheme. For the migration of a thesaurus for example, the whole could be considered as a concept scheme where several thematic groups of concepts could be designed as collections.

Labels

The SKOS model focuses on concepts therefore there is a distinction between the concept itself and the terms that may used to express this concept. Terms referring to a concept can be expressed via lexical labels according to the SKOS data model. A lexical label is a string of Unicode characters which allows you to have a term in any language with or without Latin characters. The SKOS data model defines 3 types of lexical label:

- **Preferred label**, introduced in the SKOS data model as the skos:prefLabel property, corresponds to the notion of descriptor from the standards for the elaboration of thesauri. The SKOS data model does not allow there to be more than one preferred label in the same language.
- **Alternative label**, introduced as skos:altLabel property, are mainly used to give synonyms to the preferred label or other ways to refer to this preferred label, e.g. different spellings or acronyms. The SKOS model does not forbid the exclusive use of alternative labels instead of one preferred label and many alternative labels.
- **Hidden label**, introduced by the skos:hiddenLabel property, may be used for mentioning the misspellings of preferred or alternative labels but also for mentioning obsolete forms of a term. Alternative and hidden labels correspond roughly to the USE and UF (Used For) indicators defined in the ISO standards for thesauri. By definition, hidden labels are not visible but are very useful for the retrieval. Obviously the SKOS data model does not allow the use of the same string of characters as a preferred, alternative or hidden label in the same language. An extension to the SKOS model, SKOS-XL, is proposed for modeling more precisely the labels and including morphologic or syntactic information on labels.
- **Notation**: symbols or codes that are not recognizable or understandable in any natural language. Notations are different from labels which usually are words or expressions understandable in any natural language. The skos:notation can then be used for example in the case of classifications where a code refers to a term referring itself to a concept. The notation can be more convenient than using an alternative label since it is considered as unambiguous and language independent.

The use of these different types of label enables the understanding of the concept and is useful for human-readable knowledge representation. The use of labels is not mandatory in the SKOS datamodel but is highly recommended especially for maintenance purposes.

Documentation properties

The SKOS model offers a variety of possibilities to provide information related to concepts. Different types of notes can be used to give the most accurate information. These notes can be of different natures (plain text, image, quotes ...) and be used without any restriction.

The different types of notes that can be used to document a concept are:

- Note (skos:note)
- Change note (skos:changeNote)
- Definition (skos:definition)

- Editorial note (skos:editorialNote)
- Example (skos:example)
- History note (skos:historyNote)
- Scope note (skos:scopeNote)

The skos:note can be used to provide general documentation on a concept. All the other types are specializations of this general property. The skos:changeNote and editorialNote are mainly useful for the purpose of administration and maintenance. The skos:definition, skos:example, skos:historyNote are useful for providing information on the concept for a better understanding of its meaning. As for labels, documentation properties can be provided in different languages by using language tags with the xml:lang attribute.

Semantic relations

The power of the SKOS model lies in the semantic relations that can be used to connect between different concepts. These semantic relations play a crucial role for defining concepts. There are two different categories of semantic relation:

- Hierarchical relations:

  Hierarchical relations are introduced via two properties, skos:broader and skos:narrower. The skos:broader property is used to assert that a concept has more general meaning. skos:narrower is the inverse property used to assert that a concept has a more specific meaning. One concept can have more than one broader concept or more than one narrower concept.

  It is important to note that these two properties only assert direct/immediate hierarchical link between two concepts. In order to enable non-immediate link between two concepts, the SKOS model provides two other properties that are transitive.

  As for the skos:broader and skos:narrower, the properties skos:boaderTransitive and skos:narrowerTransitive are the inverse of each other.

- Associative relations:

  The property skos:related is used to assert an associative link between two concepts. This property may be useful to make a link between a concept and another one which is neither an equivalent nor a broader/narrower concept. It is important to note that the skos:related property is symmetric.

  skos:related is not a transitive property.

  It is very important to keep in mind that, according to the guidelines provided in ISO 2788 and BS8723, mixing associative relations and hierarchical relations is not consistent with the SKOS data model. Therefore a special attention must be paid to the semantic relationships between concepts.

Mapping

The power of the SKOS datamodel relies on the mapping features it offers. The SKOS data model provides several mapping properties for making alignment between concepts from different concept schemes. These properties are:

- skos:closeMatch
- skos:exactMatch
- skos:broadMatch
- skos:narrowMatch
- skos:relatedMatch

As for semantic relations between concepts, the mapping properties can be associative or hierarchical. The skos:broadMatch and skos:narrowMatch properties are used for a hierarchical mapping link between concepts whereas the skos:relatedMatch property is used for an associative one. Exactly as for semantic relations, skos:broadMatch is the inverse property of skos:narrowMatch.

The properties skos:closeMatch and skos:exactMatch are used to make a mapping link between concepts that are very similar or equal so they can be used interchangeably. The skos:exactMatch property is transitive and symmetric. Mapping properties are used rather than semantic relations in order to make mapping links between concepts from different concept schemes. In the case of a same concept scheme semantic relationships will be used instead of mapping properties.

As for semantic relations, there may be some conflicts in mixing hierarchical mapping properties with associative ones.

Top

Guidelines for SKOSification

By SKOSification, we mean the process of conversion or transformation of a terminology into SKOS. We list below some guidelines for proceeding to this conversion from a technical and organisation point of view. From the technical point of view, many of the guidelines provided here are inherent to the SKOS model but a special attention must be paid to these points in order to enable the general consistency with the netwok of terminologies.

Evaluate the main features of the terminology to be migrated

Before starting any procedure for converting a terminology into SKOS, the institution must have defined the purpose of its terminology (e.g. indexing and retrieval, only indexing, or only retrieval). As a second step, and a consequence of the definition of the purpose, the institution must evaluate if SKOS is the appropriate format considering the content of its terminology. In the case of authority files for instance, SKOS may not be the most appropriate format. Here are some features that can help for this evaluation:

- *Concepts*: Is the terminology dealing with objects and abstract things that could be assimilated to concepts? Is the terminology dealing with persons? => if the terminology is dealing with persons and not objects or abstract things, a standard like FOAF (Friend Of A Friend)FOAF: http://www.foaf-project.org would be more apropriate
- *Semantic relations*: Are the descriptors (then concepts) of the terminology can be linked together via semantic relations. => if the terminology only contain independent descriptors without any semantic relations, a SKOS modelization is not absolutely necessary, an RDF representation may be more convenient.
- *Interoperability*: Can the terminology be linked to another resource dealing with the same subject/domain or scope? => if the terminology can be linked to other resources, all the potential links should be considered before the transformation process in order to implement these links in a most efficient way.

Identify your concepts

Interoperability: Can the terminology be linked to another resource dealing with the same subject/domain or scope? => if the terminology can be linked to other resources, all the potential links should be considered before the transformation process in order to implement these links in a most efficient way.

- Use of a Persistent Identifying System for the definition of the URIs

  As we described them above, we recommend the use of standards for the identification of the concepts. Indeed, as the identification of concepts is achieved with the definition of HTTP URIs, these URI must be declared to persistent identification systems such as PURL which is normalised. This will also be of a great benefit since it is location-independent, e.g. if the terminology is moved from one location (housing server) to another, the URIs identifying the concepts of this terminology will not have to be modified.

- Use of non-explicit URIs

  It is highly recommended to use non-explicit URIs in order to avoid the reuse of a same URI for identifying two different concepts. Indeed as natural languages are by definition

ambiguous and polysemous, it is possible that two different concepts might have two similar labels. The use of explicit URIs supposes that the choice of one specific natural language has been made during the definition or the migration of the terminology which cannot be convenient in a multilingual context.

See the booklet on Persistent identifiers: recommendations (PDF).

Define with precision the labels expressing concepts

- Preferred labels must be unique within a concept scheme

  As it is required by the SKOS data model, no two concepts from a same concept scheme should have the same preferred label in a given language. However as natural languages are highly polysemous and full of homographs, the SKOS data model does not forbid that one concept can have two same preferred labels in two different languages.
  Each concept must be expressed with one preferred label per language (mandatory)

  As we saw above, the SKOS data model does not forbid the absence of preferred label, but labels are meant to help the understanding and refining the meaning of a concept. This is especially true in a multilingual context and it is helpful for purposes of administration and maintenance. Therefore we recommend using one preferred label per language. It is important to note that this also means that is not possible to have several preferred labels in the same language.
  Avoid the concatenation of several words for a same label

  In order to get the most accurate description, we recommend avoiding several values as a preferred term. For example, double concepts such as “dwelling/houses” must be considered as two different concepts that are linked by a semantic relation. The use of scope notes can help to reinforce the closeness of these two concepts. The link between the two terms must be defined in order to provide the best description. We can state that “dwelling” and “houses” are synonyms; then the double concepts can be modelled as follows: Dwelling: preferred label and houses: alternative label

  Another possibility in the case of double concepts is to model the two concepts as related concepts.

- Privilege the use of the lemma for the preferred label and possibly the other labels

  The preferred label should consist in a single word term or a compound words term in natural language. This means that no artificial word or code must be used to label a concept. Such code must be defined using the skos:notation property. The lemma of a word represents its canonical form. We strongly recommend this form of terms to be used as preferred label. For instance, in English or in French, the usual form of a lemma in the case of nouns is the singular for the number and the masculine for the gender.
  Privilege the typography in use by convention in the languages involved

  The labels should respect the typographical rules that are usually in use in the languages of the labels. For instance, in English all the words referring to a language or nationality starts with an upper-case character whereas in French, these words will be in lower case characters. Thus we recommend respecting the conventions that are in use for each language involved. Any exception to this guideline must be documented via documentation properties of the model.

  For verbal forms, infinitive forms will be privileged. Thus the forms of terms should be based on the conventions in the languages involved. If the concept is only expressed with labels in specific forms that do not correspond to the lemma, this must be documented via the documentation properties (skos:note, skos:changeNote, skos:editorialNote or skos:historyNote) In the case of compound terms, if possible, the addition of adjectives or verbs to a noun phrase should be limited. In the same spirit, the use of articles and prepositions should be avoided in order not to extend the length of the label. From the computing systems point of view, these guidelines can help the efficiency of a retrieval system.

- Avoid the duplication of information

  The SKOS data model consists of classes and properties as we saw above. Meanings are to be deduced by an efficient use of these properties. As some of the properties

available in the SKOS model are proposed as pairs (inverse or symmetric), this supposes that the use of one property implies the opposite or the reverse. Therefore it is better to avoid duplication and not to repeat the same information in different ways. SKOS terminologies are processed by machines. So the less redundant information there is, the faster the results of a query can be retrieved. The main properties to pay attention to in order to avoid duplication of information are:

- Inverse properties

  The use of the skos:broader or skos:narrower property implies the inverse meaning. Asserting that A has a broader concept B implies that B has a narrower concept A. This is true also for the skos:broaderTransitive and skos:narrowerTransitive property.

- Symmetric properties

  The skos:related property is symmetric then if an assertion that A is related to B is made, there is no need to make the following assertion, B is related to A.

Provide precision to the semantic relations of your concepts

- Non-immediate hierarchical relations

  In some cases, semantic relations between concepts have to be described with precision in order to avoid a loss of meaning or information and also avoid designing information which will not make any sense. For example the skos:broaderTransitive/skos:narrowerTransitive pair of properties allows to describe with precision relations between concepts when two levels of hierarchy are impacted. Then the use of these transitive properties is preferred in order to assert a non-immediate hierarchical relationship between two concepts. However there is a possibility to use an extension to the SKOS data model in order to remove the symmetry of a property if this creates confusion in the meaning of the concepts.

- Consistency of the semantic relations

  In order to ensure consistency, mixing hierarchical relationships with associative ones should be avoided. For example, a concept A cannot be related to another concept B if this concept A is the narrower concept of a concept C. Therefore a special attention must be paid when designing the semantic relations between concepts.

Ensure the documentation of concepts and the terminology

- Provide documentation for each change that may occur to a concept and its labels

  The SKOS data model provides number of documentation properties in order to refine the meaning of a concept or keep track of the changes on the label(s) of a concept and/or its meaning. For the purposes of administration and maintenance of the terminology, each change must be reported in the SKOSified terminology using change notes (skos:changeNote) or editorial notes (skos:editorialNote).

- Provide as much as possible documentation to concepts with scope notes

  As mentioned above, documentation on concepts helps to refine the meaning of a concept. The use of scope notes (skos:scopeNote) can be very helpful in enabling a better understanding of the concepts with contextual information. Examples may also be provided via skos:example property. Documentation of concepts is especially needed in the case of homographs/homonyms in the same language or different languages for the labels expressing the concept. Then scope notes and examples can provide the user with a semantic disambiguation.

Top

Guidelines for mapping

Mapping is an inherent part of the SKOSification of a terminology. The following guidelines emphasize some aspects of the mapping process that may be crucial for general consistency of the terminology and the meanings of concepts.

Pay attention to the identification of your concepts during the mapping process

- Use only absolute URIs

  This guideline follows on from the one referring to the identification of concepts in the SKOSifcation part above. The terminology is made available in a machine-readable format by the SKOSification process. In order to make easily computable the identification of concepts and linking between concepts, it is recommended to use absolute URIs rather than relative ones.

  For example:

  <rdf:Description rdf:about="http://www.athenaeurope.org/athenawiki/AthenaThesaurus/RMCA_Keywords#architecture"> is an absolute HTTP URI

  <rdf:Description rdf:about="RMCA_Keywords#architecture"> is a relative HTTP URI.

- Respect the URIs of the original sources

  As URIs are defined in order to identify the concepts uniquely, during the mapping process from a concept scheme to another, the URI defined within each concept scheme must be respected in order to enable the interoperability between the different resources involved.

Avoid the duplication of information

We saw that the structural properties for defining the semantic relations between concepts are either inverse or symmetric. This is also true for the mapping properties.

- Inverse properties

  The mapping properties skos:broadMatch and skos:narrowMatch are each other's inverse therefore there is no need to repeat twice the same mapping link using both properties for the same subject and object.

- Symmetric properties

  The mapping property skos:exactMatch and skos:closeMatch are symmetric. So repeating the mapping link can be avoided. The property skos:exactMatch is also a transitive property then there is no need to repeat the mapping link on several levels.

  For instance: A skos:exactMatch B B skos: exactMatch C

  The assertion A skos:exactMatch C can be inferred from the preceding statement.

Provide precision to the semantic relations of your concepts

- Use the appropriate properties to make links between concepts

  The SKOS data model provides semantic relations and mapping properties, and does not restrict the use of these properties. However we strongly recommend to model in a homogenous way the relations between concepts in order to ensure the semantic consistency of the terminology. We recommend to:

  - Use mapping properties to make a link between concepts from different concept schemes
  - Use semantic relations properties to make a link between concepts within a same concept scheme

The SKOS data model does not forbid using semantic relations properties for making a link between concepts from different concept schemes but it is highly recommended to follow these guidelines.

Enable the multilingualism

- Manage multilingualism of the terminology through mapping of concepts and terms

  The mapping process can be useful in a monolingual context but is especially relevant in a multilingual context. Equivalences can be stated from the mapping links made between several terminologies in different languages. Equivalencies in a multilingual context can be of three kinds: semantic, cultural or structural. The semantic aspect refers to the meaning of the concept; the cultural aspect refers to the use of a term in a given language or culture; and the structural aspect refers to the semantic relations between concepts. This last aspect deals with the mapping and allows defining complete equivalence (synonymy) or partial equivalence (quasi synonymy) or nonequivalence. As it was the case for the first version of the ATHENA Thesaurus, equivalences between concepts in languages that were not initially involved in the source terminology can be deduced from correct mapping links without translating the concepts.

Ensure the documentation of concepts and the terminology

- Make explicit with notes the purpose of a relation

  For the purposes of maintenance and administration, it is important to explain the choices of modelling that have been made for making links between concepts. The use of scope notes can help making explicit these choices. Documentation properties can also keep track of history of mapping links. Validation is an important part of the SKOSification process and mapping also. Therefore a special attention must be paid to this final step of the SKOSification. From a technical point of view, in order to check the consistency of your converted terminology to the SKOS model, we recommend using the online web service [Party]. Pool Party offers a free online tool for validating SKOS files that may be already online or stored on your local repositories. This tool checks the consistency of the SKOSified terminology according to the following points which refer to our guidelines:

  - Valid URIs: the tool checks if there is not any unauthorised character in the URI. Although if an URI is used twice for identifying two different concepts, there won't be any alert or warning.
  - Missing language tags: the tool checks if all the labels and notes have a language tag
  - Missing labels: the tool checks that each concept has at least one preferred label.
  - Loose concepts: all the concepts that are isolated and not linked to other concepts are pointed out as loose concepts
  - Disjoint OWL classes: the tool checks the eventual consistency with OWL elements that may be in the SKOSified terminology
  - Consistent use of labels: the rules for the use of labels are checked by the tool in order to avoid the use of a same label as a preferred label and alternative or hidden label, and to avoid the use of two preferred labels in a same language, …
  - Consistent usage of mapping properties: the tool checks the consistency in the mapping relations.
  - Consistent usage of semantic relations: the tool checks that there is no mix between hierarchical and associative semantic relationships.

  From the content point of view, only the administrators and users of the terminology can validate the final migration of the terminology into SKOS format at least for an initial transformation process since they will be the one able to confirm or modify the general design of the terminology and its semantic relations according to the indexing and retrieval efficiency. For further modifications and updates, a set of rules and policies have to be defined in order to enable the collaborative moderation for managing the terminology. These rules and policies have to be agreed on by the community of users.

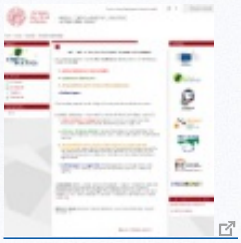
Top

Linked Heritage LEARNINGObjects

TERMINOLOGY

TERMINOLOGY MANAGEMENT PLATFORM (TMP)

Readings and training materials



Terminology Management Platform (TMP)

TMP stands for Terminology Management Platform. The TMP demonstrator, presented here and available on www.culture-terminology.org, provides the functionalities identified within the project as needed to share a terminology on the basis of the 6 functional steps identified for terminology management: Registration, SKOSification, Enrichment, Mapping, Search and Navigation, Collaboration. The demonstrator allows cultural heritage institutions to experiment with transforming their in-house terminologies to the SKOS format, to deal with metadata registration and version management of their terminologies, to map and enrich these terminologies according to the principles of the Semantic Web etc.

Content sections:

Top

Getting started

The TMP counts 4 user roles that are allowed to make some specific actions.

- **General User:**
  The general user is the visitor who could browse and navigate through the terminology registry. The general user can search into the registry and view terminology that are registered as public ones. The general user is not logged in.

- **Contributor:**
  The contributor is allowed to view public thesaurus and make suggestions to modify them. The contributor is a user who has an account on TMP and is logged in. The contributor may belong to another institution and is then able to make suggestions to a public terminology.

- **Editor:**
  The editor works under the authority of the owner of the terminology. The editor has the same rights than a contributor but can also have some management rights on his terminologies. The editor is then able to Create, Read, Update and Delete concepts and terms.

- **Owner (admin):**
  The owner has a role of administrator of terminologies; the owner is responsible for all the terminologies of his institution. So he can Create, Read, Update and Delete concepts and terms and also accept or refuse the suggestions made by the other users. The owner is also the one responsible for managing rights for the users and editing the metadata of the terminologies. He is the one able to publish the terminology and set it as private or public.

These user types are not exclusive, a same individual can be each of this user depending on the context and the terminologies being used.

In order to have access to the main features of the TMP, the user will have to be logged in to the TMP. If the user already have an user account on the MINT tool then he will be able to use its login and password to connec to TMP (Login).
The user will need to sign in if he does not have an access to MINT or if he wants to connect to the TMP for the first time.

Top

Search terminologies

[Click to enlarge]

There is no need to be logged in to search for terminologies that are published. For each terminology the Title, the subject(s), language(s) and the author is visible. It is possible to sort each of these fields and then view the terminology.

When the user is logged in to the TMP, he can search through his own terminologies or through all the public terminologies.

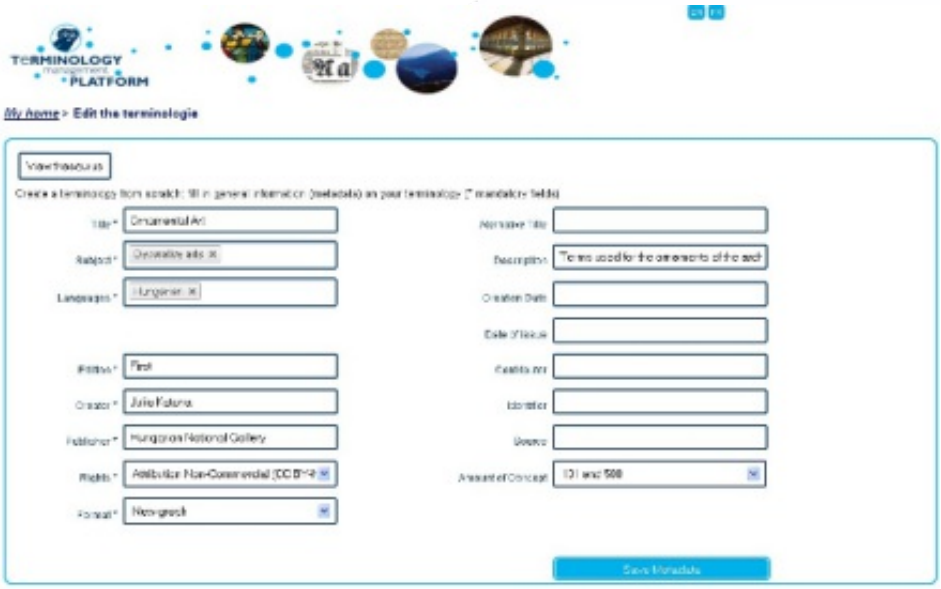
[Click to enlarge]

Top

Create new terminologies

The first step to create a new terminology is to provide some metadata on this terminology.

These metadata fields have been selected from the Dublin Core description scheme. The mandatory fields are marked with a * sign.



[Click to enlarge]

The owner of a terminology will be the one responsible for providing or updating these metadata.

In the case of the creation of a terminology, its format will be “new graph” instead of CSV or SKOS.

Top

Import a terminology (SKOSification)

The import of a terminology stands for the import of an already SKOSified terminology and a terminology that still have to be SKOSified as well.

[Click to enlarge]

For the time being only two input formats are available, CSV and SKOS formats.

In the case an institution only has a spreadsheet file, the spreadsheet must be adapted in order to be uploaded in the TMP and to proceed to the mapping to SKOS. Attention must be paid to the format of the spreadsheet only the standard format CSV is accepted so the terminology must be saved as a CSV file if it has an .XLS or .ODS extension.

For example an excel file which is formatted like in the figure below will have to be formatted according to a specific methodology.

	C	D	E	F
1	prefLabel@nl	prefLabel@eng	broader	SN@nl
2	Objectnaam	Object name		
3	Architectuur	Architecture	41875	Topterm voor architecturale elementen en onderdelen
4	Bouwwerk	Single built work	41877	Algemene benaming voor alle soorten van (in aanbouw
5	Deel van een gebouw	Building division	41879	Wordt gebruikt voor het benoemen van afzonderlijke k
6	Funerair bouwwerk	Funerary structure	41879	Wordt gebruikt voor gebouwen die betrekking hebben
7	Monument	Monument	41879	Bouwwerken die getuigen van een (vroegere) cultuur,
8	Architecturaal element	Architectural element	41877	Structurele of decoratieve vormen die oorspronkelijk o
9	Antefix	Antefix	41878	Dakornamenten, die het open eind van een rij dekteg
10	Boog en boogonderdeel	Arch and arch component	41878	Gidsterm voor bogen en onderdelen van bogen
11	Korbeel	Bracket (structural element)	41878	Wordt gebruikt voor architecturale elementen die dien
12	Haard en haardonderdelen	Hearth and hearth component	41878	Gidsterm voor haarden en onderdelen van haarden
13	Zuil en zuilonderdeel	Column and column component	41878	Gidsterm voor zuilen en onderdelen van zuilen
14	Onderdeel van een pyramide	Pyramid component	41878	Wordt gebruikt voor onderdelen van pyramiden
15	Kraagsteen	Corbel	41878	Wordt gebruikt voor een uit de muur vooruitspringende
16	Oppervlakte-element en onderdeel va	Surface element and surface element	41878	Gidsterm voor architecturale elementen die een deel v
17	Entablement en entablementonderdel	Entablature and entablature component	41878	Wordt gebruikt voor de omlijsting van een raam of deu

[Click to enlarge]

In order to be parsed correctly by the SKOSification module of the TMP, this Excel file will have to be exported as a CSV file. This could be done quite easily just clicking on Menu “File”> Save as. Then the the Excel file will be exported as a CSV file.

In order to be parsed correctly by the SKOSification module of the TMP, this Excel file will have to be exported as a CSV file. This could be done quite easily just clicking on Menu “File” » Save as. Then the the Excel file will be exported as a CSV file.

Once the file is exported in CSV, you can see how it looks like in the following figure:

```
"ID"    "prefLabel@fre" "prefLabel@nl"  "prefLabel@eng" "broader"       "SN@nld"        "SN@eng"        "SN@fre"
41875   "Nom de l'objet"        "Objectnaam"    "Object name"
41877   "Architecture"  "Architectuur"  "Architecture"  41875   "Topterm voor architecturale elementen en onderdele
n van architectonische elementen"
41879   "Construction"  "Bouwwerk"      "Single built work"     41877   "Algemene benaming voor alle soorten van (i
n aanbouw zijnde) gebouwen"
45278   "Composant de bâtiment" "Deel van een gebouw"   "Building division"     41879   "Wordt gebruikt voor het be
noemen van afzonderlijke kamers of ruimten van een gebouw"
46319   "Construction funéraire"        "Funeralr bouwwerk"     "Funerary structure"    41879   "Wordt gebruikt voo
r gebouwen die betrekking hebben op begravingen"
41134   "Monument"      "Monument"      "Monument"      41879   "Bouwwerken die getuigen van een (vroegere) cultuur
, kunst, stijverheid of wetenschap en die beschouwd worden als uitzonderlijk cultuurhistorisch waardevol"

41876   "Élément d'architecture"        "Architecturaal element"        "Architectural element" 41877   "Structurel
e of decoratieve vormen die oorspronkelijk of hoofdzakelijk ontstonden als architectonische componenten. Deze kunne
n een structurele of een decoratieve functie hebben"
45650   " Antéfixe"     "Antefix"       "Antefix"       41878   "Dekornamenten, die het open eind van een rij dekte
gels afsluiten of verbergen, o.a. om te voorkomen dat regen tussen de daktegels kan doorsijpelen en wind de bovenst
e tegels niet kan losmaken"     "Plaque closing the outer end of the final cover tile in each row of overlapping
cover tiles running down from the ridge to the eaves of a sloped roof on Classical Greek and Roman and Neo-Classica
l buildings"
```

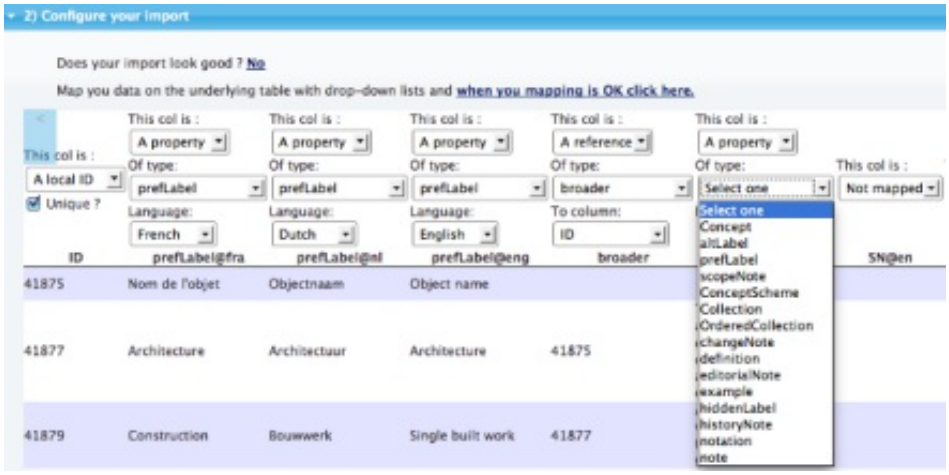
[Click to enlarge]

The import process can start with uploading the terminology from personal files. A control is done at this point to check whether a CSV file has been uploaded in conformity with what was declared during the metadata filling in.

[Click to enlarge]

Each column should show an element from the terminology. If the content from the CSV file is not well displayed, the choice of the separator must be revised according to the separator you have chosen for exporting your Excel file as a CSV file. These parameters can be changed answering the question right under the “Configure your import” tab, “Does your import look good”, click on “no” to change the parameters.

The dialog box above will show up. Once the parameters have been correctly changed, the imported csv file should display each concept as a line and each SKOS property as a column.



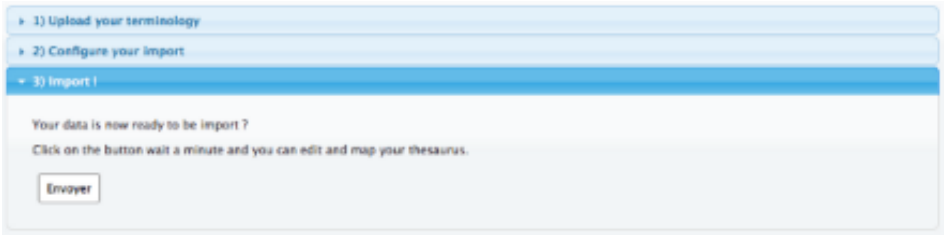
[Click to enlarge]

Now that the import is well formed you can proceed with the mapping of your input terminology with the SKOS properties.

Each column can be a local ID, a property or a reference. In the case of a property, the common skos:properties will be suggested, for example: a concept, labels (preferred, alternative or hidden), notes (scope note, editorial or history note,...). For each property, the language attribute is mandatory. By default, the languages that have been declared during the registration of the terminology in the metadata will be proposed to the user. In the case of a reference, all the properties for SKOS relationships will be suggested, for example: broader, narrower, related; mapping properties such as exactMatch or closeMatch will be

also suggested. This mapping completely depends on the structure of the input CSV file.

When the mapping is done then the import process can start and the SKOSification will be done simultaneously.



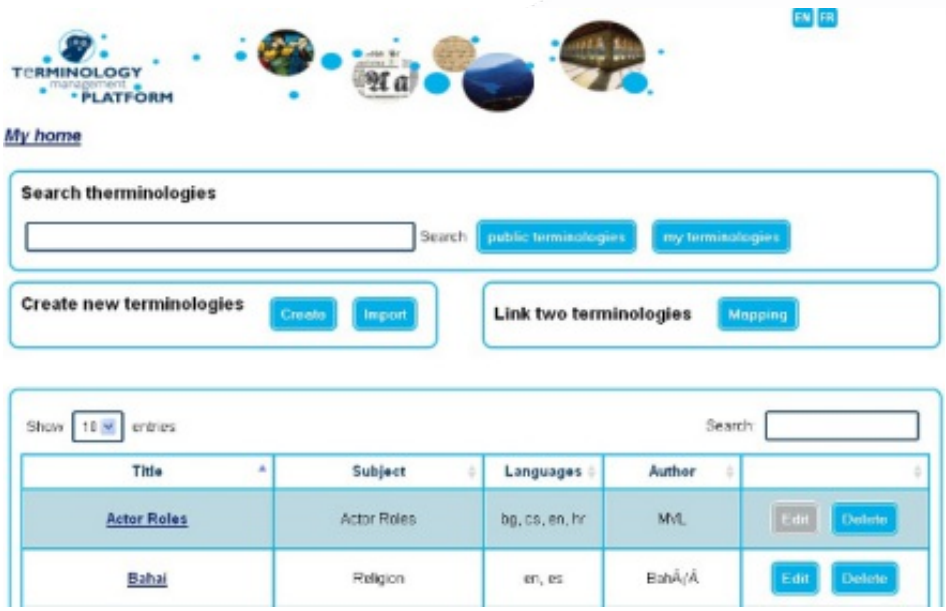
[Click to enlarge]

If the terminology is already available in SKOS, it will be directly uploaded with a syntax and validity checking. In conformity with the SKOS datamodel, each concept must have one and only one preferred label per language. Each property must have a language attribute.

Top

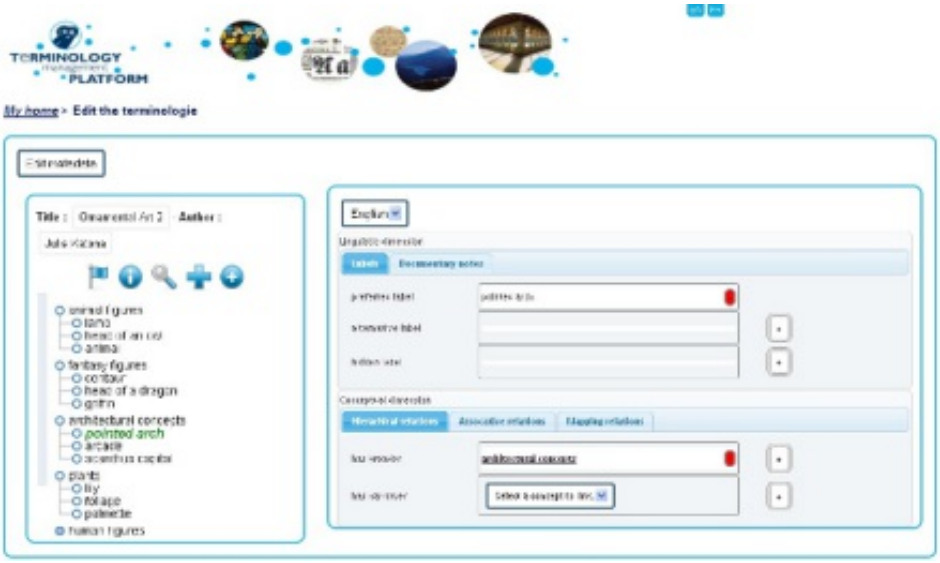
Edit a terminology

If the terminology is already available in SKOS, it will be directly uploaded with a syntax and validity checking. In conformity with the SKOS datamodel, each concept must have one and only one preferred label per language. Each property must have a language attribute.



[Click to enlarge]

The user will access to the « edit » button when browsing his terminologies. The Graphical User Interface has been designed in order to be very intuitive and easy to understand for the users.



[Click to enlarge]

The user will access to the « edit » button when browsing his terminologies. The Graphical User Interface has been designed in order to be very intuitive and easy to understand for the users.

	Move up or down the concept in the tree of concepts
	Delete the concept

When the concept is selected, the selected concept is then displayed in green color and the related information is displayed on the right with two specific dimensions. Indeed a clear separation is done between the conceptual and the linguistic dimensions:

- The linguistic dimension of the terminology in a dedicated frame for the SKOS linguistic relationships i.e. the labels (skos:prefLabel, skos:altLabel, skos:hiddenLabel) and the documentary notes (skos:note, skos:scopeNote, skos:definition, skos:example, skos:historyNote, skos:editorialNote, skos:changeNote).
- The conceptual dimension of the terminology in a dedicated frame for the SKOS semantic relationships e.g. the hierarchical relationships (broader and narrower) and the associative relationship (related). The mapping relations resulting from the mapping procedure will be displayed on the dedicated tab.

The flag icon on the left block will allow the user changing the language of the tree of concepts (according to the languages that were declared in the metadata) but the language can also be changed on the right block. This distinction between the language of the tree of concepts and the one for editing/enriching the selected concept is foreseen in order to enable the multilingualism of the terminology. Such a feature allows defining labels in a different language from the language used for ontology building Regarding the conceptual dimension, the TMP logical properties define useful guidelines for terminology building. They are integrated into the TMP interface. For example, the irreflexivity and acyclic properties of the SKOS semantic relationships allows proposing to user only the concepts that verify these properties.

The edit functionality can also be called from the mapping interface.

Top

Link two terminologies (mapping)

The SKOS data model provides several mapping properties for making alignment between concepts from different concept schemes. Within TMP, we can consider that each terminology from an institution is SKOS:ConceptScheme. The SKOS mapping properties are the following ones:

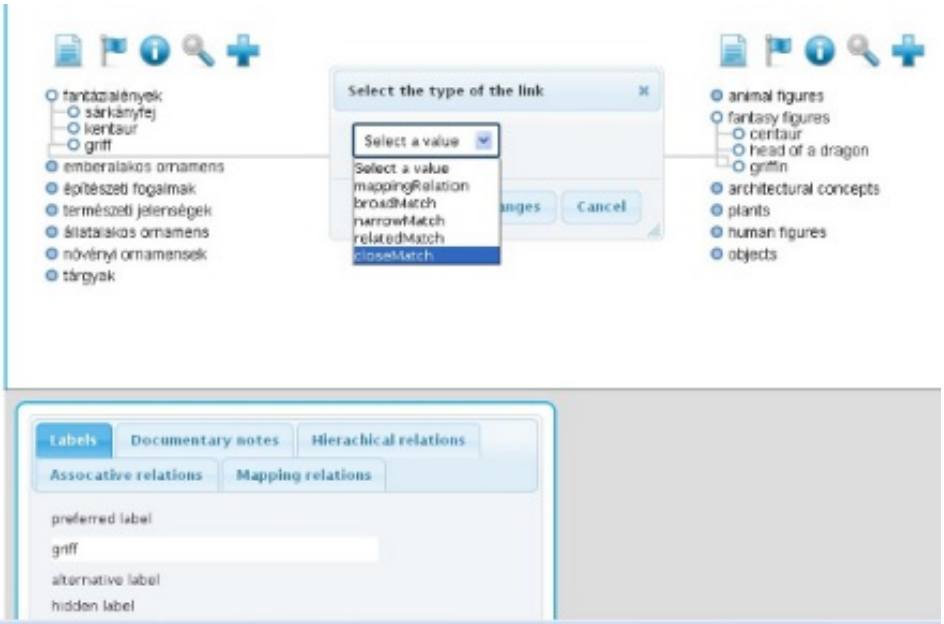
- skos:closeMatch

- skos:exactMatch
- skos:broadMatch
- skos:narrowMatch
- skos:relatedMatch

As for semantic relations between concepts, the mapping properties can be associative or hierarchical. The skos:broadMatch and skos:narrowMatch properties are used for a hierarchical mapping link between concepts whereas the skos:relatedMatch property is used for an associative one. Exactly as for semantic relations, skos:broadMatch is the inverse property of skos:narrowMatch.

The properties skos:closeMatch and skos:exactMatch are used to make a mapping link between concepts that are very similar or equal so they can be used interchangeably. The skos:exactMatch property is transitive and symmetric. Mapping properties are used rather than semantic relations in order to make mapping links between concepts from different concept schemes. In the case of a same concept scheme semantic relationships will be used instead of mapping properties.
As for semantic relations, there may be some conflicts in mixing hierarchical mapping properties with associative ones.

TMP provides a graphical tool to for the mapping process. The graphical user interface has been defined in order to be the most intuitive possible since the concepts can be mapped together easily with a simple drag and drop and the type of mapping can also be specified.



[Click to enlarge]

The figure above illustrates the mapping tool of the TMP. The terminology of the institution will be displayed on the left and the terminology to be mapped to will be displayed on the right. The title and the author of the terminologies to be mapped are displayed on the top of the columns. The icons right under the title of the terminology can do the following actions:

	Upload a terminology for mapping
	Change the language of the terminology (the languages that were declared in the metadata form will be suggested to the user)
	See the metadata of the terminology
	Search for a concept in the terminology
	Add a new concept

When a concept is selected, a block displaying the details on the selected concept appears on the bottom of the column. The information on the concept that is displayed in that block is separated in different tabs for the labels, the documentary notes, hierarchical relations, associative relations and mapping relations. We will see in the edit part that each concept is displayed in this way. This information on the selected concept can help the user during the mapping process as it can reduce the ambiguity or guide the user to define the type of link that will be chosen.
The procedure to make the link is quite easy since the user must select the concept in his terminology (left column) and the concept to be mapped with in the other terminology (right column) and then a pop up box will suggest the type of link that must be made between the two selected concepts.

Once the mapping is done the link between the two concepts appears in grey. One upcoming development is to distinguish the different mapping links using different colors.

Top

« Back Forward »

Linked Heritage LEARNINGObjects

TERMINOLOGY

LITERATURE AND GLOSSARY

Readings and training materials

Literature

Shiri A., *Powering search. The Role of Thesauri in New Information Environments*, New Jersey 2012

Aitchison J., Gilchrist A., Bawden D., *Thesaurus Construction and use: A Practical Manual*, 4th. ed. London 2000

Hodge G., *Systems of knowledge organization for digital libraries: Beyond traditional authority files*, Washington 2012

ISO 25964-2, *Information and documentation - Thesauri and interoperability with other vocabularies* – Part 2: Interoperability with other vocabularies, Geneva, March 2013

BS 8723: Structured Vocabularies for Information Retrieval, London 2005

ANSI/NISO Z39: *Guidelines for the construction, format, and management of monolingual controlled vocabularies*, Bethesda, 2005

SKOS: http://www.w3.org/2004/02/skos/

Glossary

Concept:

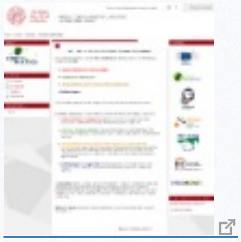
- Controlled vocabulary
- Information retrieval
- Interoperability
- ISO-norms
- Linked Data
- Mapping
- Thesauri
- URI
- URL
- Query
- RDF
- SKOS
- Standards
- XML
- Semantic Web

Linked Heritage LEARNINGObjects

TERMINOLOGY

3/4 – CONCLUSIONS

Readings and training materials



Thank you for viewing our Learning Object *Terminology*

We invite you now to answer some questions for your learning assessment and to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: Terminology.

For further details:

Readings and training materials

« Back Forward »

Linked Heritage LEARNINGObjects

TERMINOLOGY

4/4 – CREDITS

Project by:

Graphic and Web design by Gianluca Drago
CAB, University of Padua (Italy)

Funded by the Competitiveness and Innovation Framework Programme, 2011.

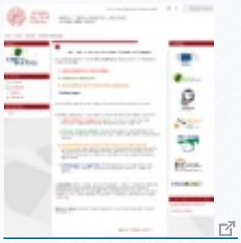
« Back

Linked Heritage LEARNINGObjects

LINKING CULTURAL HERITAGE INFORMATION

ABOUT THE LEARNING OBJECT

Readings and training materials



Summary

This LO focuses on the current structure of the Linked data Cloud - the best known representation of linked data. Size, licenses, subjects... all the informations were extracted from the analysis made by *Linking cultural heritage information*, the 2nd Work package of *Linked Heritage*. At the end, Cultural heritage institutions dealing with linked data publication (museums, libraries, archives..) can find best practice recommendations.

Target audience

- Decision makers and project managers of public libraries, museums, archives.
- Culture and information specialists from museums, libraries and archives.

Thanks to Linked data – especially Linked open data – they can enrich the knowledge and improve their visibility on the web.

Forward »

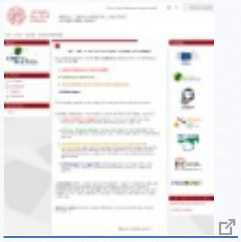
Linked Heritage
LEARNINGObjects

LINKING CULTURAL HERITAGE INFORMATION

1/5 – WHY LINKED DATA?

Texts and data are mostly taken from McKenna, G. *Linked heritage experience in linking heritage information*. JLIS.it. Vol. 4, n. 1 (January 2013): Art: #6304. DOI: 10.4403/jlis.it-6304. Web.

Readings and training materials



Linked data is a set of best practices required for publishing and connecting structured data on the web for use by a machine.

When this is accomplished in an open manner, whereby everyone can access, reuse, enrich and share the data published, we talk about *Linked Open Data* – as a technology used to realize the Semantic web.

So we can say that:

- data can be open but not linked;
- data can be linked but not open;
- the Sematic web can only function with data which is both open and linked.

Publishing linked open data online and making connections with other data of the same type allow content providers to gain knowledge, improve their visibility on the web and obtain multiple benefits:

- Reduced duplication of information;
- Constant data updating;
- Higher profile with more traffic to their websites;
- Authority;
- New audience;
- Better user experience;
- Efficient use and reuse of resources.

> *"We can use the Mona Lisa as an example. If one provider describes the Mona Lisa as a painting by the Italian artist Leonardo Da Vinci and another source states it is a painting of a woman, linking these pieces of data will allow people to find the Mona Lisa by searching for paintings of women by Italian artists."* [1]

Linked Open Data has a key role in the Europeana Strategic Plan 2011-2015 and supports the idea that Europeana distributes content and engages its users.

NOTES

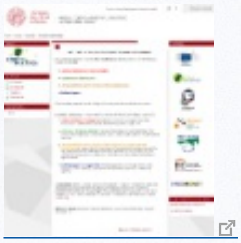
1 Maarten Zeinstra and Paul Keller. *Open Linked Data and Europeana*, (PDF)

Linked Heritage LEARNINGObjects

LINKING CULTURAL HERITAGE INFORMATION

2/5 – ANALYSIS OF THE CLOUD BY LINKING CULTURAL HERITAGE INFORMATION

Readings and training materials



Page Index

Linking cultural heritage information is the name of the 2nd Work package (WP2) of *Linked Heritage* – a 30 month EU project, started on 1st April 2011 and coordinated by ICCU (Rome) - Istituto Centrale per il Catalogo Unico delle biblioteche italiane.

Some of the objectives set by WP2 are:

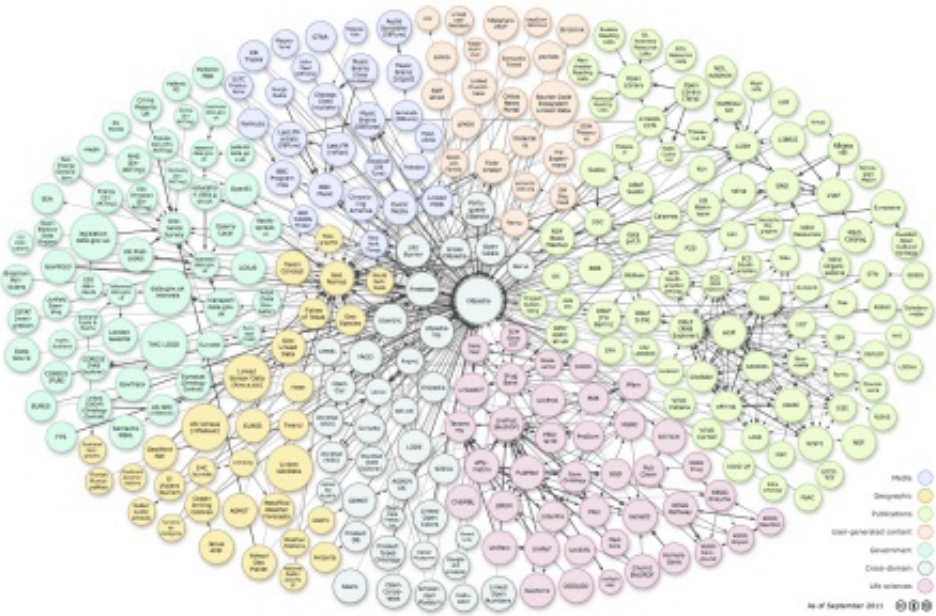
- to explore the state of the art in linked data and its applications and potential;
- to identify the most appropriate models, processes and technologies for the deployment of cultural heritage information repositories as linked data.

In particular, WP2 examined both the structure and information carried by the Linked data Cloud.

Top

The Cloud

The Cloud is the best known representation of linked data. It shows “packages” of linked data and the links between packages. It is growing very quickly and the most recent diagram (September 2011) counts 331 packages.



The Cloud in September 2011. [Click to enlarge].
The Linking Open Data cloud diagram is maintained by Richard Cyganiak (DERI, NUI Galway) and Anja Jentzsch (HPI).
The cloud is maintained on The Data Hub website, a W3C project which represents a registry of open and not-open knowledge, with information on packages and projects.

THE DATA HUB

The Data Hub website is managed by Linking Open Data community project – which is part of the W3C's Semantic Web Education and Outreach Interest Group (SweoIG). Therefore it may be considered as representing a significant proportion of the linked data available.

For each package *The Data Hub* website gives information about:

- name and description;
- links to the resources available;
- intellectual property rights status;
- which other packages are linked to (including number of links);
- the number of "triples" in the package (a measure of size);
- subject information and the formats used.

From WP2 analysis has emerged an unexpected image of The Cloud, beyond guiding principles and plans.

Top

Is The Cloud open?

Given that in The Cloud *open* means able to be reused, enriched and shared (also commercially) the analysis shows that a significant component of The Cloud is not open.

IPR Status	% by Package
Open	**42.6**
Not open	**57.4**

One reason for this anomaly may be that the first packages published in The Cloud didn't

consider important having a licence.

Top

Which IPR licences are used?

Open licences

Of the 132 packages with open licences:

Licence type	% by Package
Creative Commons Attribution (CC BY)	28.8
Creative Commons Attribution Share Alike (CC BY-SA)	18.2
Open Data Commons Public Domain Dedication and Licence (ODC PDDL)	10.6
Creative Commons CC Zero (CC0)	9.1
UK Crown Copyright with data.gov.uk rights	7.6
Other (Public Domain)	6.8
Other (Open)	5.3
Others	12.9

CC0 is a relatively new option and it is the choice made by Europeana – and secondly by its providers – for its publication of linked open data. It is the most permissive of the open licences, with attribution being a recommendation rather than mandatory.

Not open licences

Of the 178 packages with licences that are not open, or with no licence information:

Licence type	% by Package
not given	69.1
None	14.6
Creative Commons Attribution Non-commercial (CC BY-NC)	7.3
Other (Not Open)	6.7
Creative Commons Attribution (CC BY)	1.1
Other (Non-Commercial)	0.6
Creative Commons Attribution Share alike (CC BY-SA)	0.6

For over 80% of packages of this part of The Cloud there is no information about the IPRs: a

large part of published linked data does not seem to have a licence for its use. The result is that it is unclear what can be done with this data.

Top

How big is The Cloud?

There are c38 billion triples in The Cloud with a large distribution in size: 9 packages (2.89%) have over a billion triples. Nearly a quarter of the packages are relatively small.

The ten largest packages with open licences are:

Package	Number of triples
LinkedGeoData	3.00 billion
UK Legislation	1.90 billion
Linked Sensor Data (Kno.e.sis)	1.73 billion
data.gov.uk Time Intervals	1.00 billion
DBpedia	1.00 billion
Open Library data mirror in the Talis Platform	0.54 billion
The Open Library	0.40 billion
Freebase	0.34 billion
transport.data.gov.uk	0.33 billion
Data Incubator: MusicBrainz	0.18 billion

The ten largest packages without open licences are:

Package	Number of triples
TWC: Linking Open Government Data	9.80 billion
Data.gov	6.40 billion
Source Code Ecosystem Linked Data	1.50 billion
2000 U.S. Census in RDF (rdfabout.com)	1.00 billion
PubMed	0.80 billion
DBTune.org MySpace RDF Service	0.66 billion
UniParc	0.63 billion
DBTune.org AudioScrobbler RDF Service	0.60 billion
Linking Italian University Statistics Project	0.59 billion

UniProt UniRef	**0.49 billion**

TWC Linking open government data is the largest package in The Cloud and is an aggregation of US government data.

Top

Which are the subjects in the data?

There does not seem to be a controlled terminology for The Cloud, with the same subject represented by different tags in different packages.

WP2 analysis have combined a number of tags which appear to be the same subject. The ten most common subjects are:

Subject tag	Number of packages with tag
publications	**94**
government	**54**
life sciences	**46**
geographic	**40**
media	**32**
library	**22**
United Kingdom	**22**
education	**20**
user generated content	**19**
bibliographic	**15**

There is very little cultural heritage data. This is probably because, until the advent of Europeana, there has been no interest in linked data in this community. The appearance of *United Kingdom* as a tag shows mainly the effect of the UK Government's policy of publishing linked data. The role of the USA is not apparent, but this is because packages are not tagged *United States*.

Top

Which formats are used to encode data?

The most commonly used formats are:

Format	Number of packages using the format
Resource Description Framework (rdf)	**261**
Dublin Core (dc)	**97**
Friend of a Friend (foaf)	**84**

Simple Knowledge Organization System (skos)	57
RDF Schema (rdfs)	42
Web Ontology Language (owl)	34
Basic Geo (geo)	25
Advanced Knowledge Technologies Reference Ontology (akt)	22
eXtensible HyperText Markup Language (xhtml)	19
Bibliographic Ontology (bibo)	14
none given	13
Music Ontology (mo)	13
DBpedia Ontology (dbpedia)	12
Others	52

AKT Ontology, DBpedia Ontology, and GeoNames Ontology were developed in the context of the publication of a single package as linked data. The adoption of this type of formats by more packages suggests that these *de facto* standards are playing a significant role in The Cloud.

It is surprising, when Berners-Lee suggests using a standard format, to find that 75 formats are used by two or less packages: for the sake of interoperability it may be hoped the survival of the fittest!

Top

How is The Cloud linked?

The most important part of The Cloud is how the packages are linked together. The ten most commonly linked to packages, in terms of the number of packages linking, are:

Package being linked to	Number of packages linking
DBpedia	158
GeoNames Semantic Web	42
(none)	34
DBLP Computer Science Bibliography (RKBExplorer)	27
Association for Computing Machinery (ACM) (RKBExplorer)	26
ePrints3 Institutional Archive Collection (RKBExplorer)	26
Freebase	25
Others	72

The success of DBpedia and GeoNames is probably due to their being well-known. But the most interesting thing is that over 10% of the packages in The Cloud do not link to other packages – included in this group are some of the largest packages, *e.g.* Data.gov, 2000 U.S. Census. This shows that the linking of packages is not something that is growing in an "organic" way.

There are initiatives responsible for creating large parts of The Cloud: such an initiative would be welcome in the **cultural heritage sector** too, where Europeana is actually taking a leading role.

Top

Cultural Heritage data in The Cloud

The world of networked information is very interested in the legacy data produced by libraries, archives and museums as they are traditionally known to have a key role in producing quality information.

Unluckily there are only 18 packages in The Cloud that could be identified as having "cultural heritage" as their subject or related to it:

Package	Number of triples
VIAF: The Virtual International Authority File	200,000,000
Europeana Linked Open Data	185,000,000
British National Bibliography (BNB)	80,249,538
Hungarian National Library (NSZL) catalog	19,300,000
Amsterdam Museum as Linked Open Data in the Europeana Data Model	5,000,000
Library of Congress Subject Headings	4,151,586
Swedish Open Cultural Heritage	3,400,000
Calames	2,000,000
RAMEAU subject headings (STITCH)	1,619,918
data.bnf.fr - Bibliothèque nationale de France	1,400,000
National Diet Library of Japan subject headings	1,294,669
Gemeenschappelijke Thesaurus Audiovisuele Archieven – Common Thesaurus Audiovisual Archives	992,797
Gemeinsame Normdatei (GND)	629,582
Archives Hub Linked Data	431,088
Thesaurus for Graphic Materials (t4gm.info)	103,000
Italian Museums (LinkedOpenData.it)	49,897
Thesaurus W for Local Archives	11,000

MARC Codes List Open Data	**8,816**

The part of The Cloud from cultural heritage is still rather small (c500m triples or <1.5%), but hopefully developments from Europeana are planned to significantly increase its size. Linked Heritage project will be a significant component of it.

Top

Format

Cultural heritage packages usually follow these formats:

Format	Number of packages using the format
Resource Description Framework	**13**
Simple Knowledge Organization System	**11**
Dublin Core	**7**
eXtensible HyperText Markup Language	**4**
Friend of a Friend	**3**
Basic Geo	**1**
Bibliographic Ontology	**1**
DBpedia	**1**
Music Ontology	**1**
Object Reuse and Exchange	**1**
RDF Schema	**1**
vCard	**1**
Web Ontology Language	**1**
XML Schema	**1**

The general picture is similar to The Cloud as a whole, except that the use of SKOS is much more significant, indicating the importance of terminological resources and authority files in the sector. Of note is the absence of a format for museum information specifically. Also the Europeana Data Model is not mentioned in *The Data Hub*, according to other sources it was surely used by some packages.

Top

Link

Cultural heritage packages in The Cloud link to targets:

Package being linked to	Number of packages linking
DBpedia	**5**

Library of Congress Subject Headings	4
VIAF: The Virtual International Authority File	2
GeoNames Semantic Web	2
Dewey Decimal Classification (DDC)	2
RAMEAU subject headings (STITCH)	2
Swedish Open Cultural Heritage	1
Gemeinsame Normdatei (GND)	1
IdRef: Sudoc authority data	1
(DCMI Type Vocabulary – not in The Cloud)	1
UK Postcodes	1
AGROVOC	1
Hungarian National Library (NSZL) catalog	1
(none)	1

DBpedia and GeoNames represent well known sources of cross-domain and geographical information to link to. The rest of the linked packages are mainly other cultural heritage packages – especially standard terminologies and authority files.

Top

Serialization

RDF/XML is used by almost all the packages, but Europeana Linked Open Data uses mentions only N-Triples.

Serialisation	Number of packages using (%)
RDF/XML	16 (88.9%)
N-Triples	5 (27.8%)
Turtle	1 (5.5%)
(none given)	1 (5.5%)

This suggests that cultural heritage linked data should be, at least, published as RDF/XML and possibly as N-Triples in order to be compatible to existing data.

Top

« Back Forward »

Linked Heritage LEARNINGObjects

LINKING CULTURAL HERITAGE INFORMATION

3/5 – BEST PRACTICE RECOMMENDATIONS FOR CULTURAL HERITAGE ORGANISATIONS (MUSEUMS, LIBRARIES, ARCHIVES, GOVERNMENT AGENCIES...)

Readings and training materials

Page Index

What information to publish as linked data?

- information about all aspects of collections and their related materials (exhibition catalogues, books, history files, learning units...);
- user generated content;
- terminological resources (object types; name authorities; places; organisations...) cooperating with partner in order to avoid duplication.

Top

What licence should there be?

- Any publication of linked data must be accompanied by a licence which makes it clear what uses can be made of the data.
- The licence may be standard – like Creative Commons – or one created specifically by the publisher.

Top

How to publish the linked data?

Format standards

Established the opportunity to use basic standards – like RDF, RDFS, and OWL – it is advised to use standard formats appropriate for the type of data being published, *e.g.*

- Web resources: Dublin Core;
- Persons: Friend of a Friend;
- Terminological resources: SKOS;

- Bibliographic resources: Bibliographic Ontology;
- Music: Music Ontology.

Consider using a cultural heritage specific format for linked data, like EDM, LIDO… In particular, Linked Heritage offers the possibility of using LIDO.

How to link the packages

- link to packages of a general nature like: DBpedia; GeoNames Semantic Web; national sources of terminology;
- link to known packages in the cultural heritage, e.g.: Library of Congress Subject Headings; VIAF The Virtual International Authority File and Dewey Decimal Classification;
- provide a SPARQL endpoint to the package.

Top

EXPLORE FURTHER

Caffo, R. *"Global interoperability and linked data in libraries: ICCU international committment"*. JLIS.it. Vol. 4, n. 1 (January 2013): Art: #8726. DOI: 10.4403/jlis.it-8726. Web.

Guerrini, M., T. Possemato. *"Linked data: a new alphabet for the semantic web"*. JLIS.it. Vol. 4, n. 1 (January 2013): Art: #6305. DOI: 10.4403/jlis.it-6305. Web.

Maarten Zeinstra, M. and Kennislan, P. K. *"Open Linked Data and Europeana"* (September 2013)
http://pro.europeana.eu/c/document_library/get_file?uuid=374c381f-a48b-4cf0-bbde-172cf03672a2&groupId=10602

Europeana professionals – Linked Open Data (September 2013)
http://pro.europeana.eu/web/guest/linked-open-data

Paul walk's weblog. *"Linked, open, semantic?"* (September 2013)
http://blog.paulwalk.net/2009/11/11/linked-open-semantic

Top

« Back Forward »

Linked Heritage
LEARNINGObjects

LINKING CULTURAL HERITAGE INFORMATION

4/5 – CONCLUSIONS

Thank you for viewing our Learning Object *Linking Cultural Heritage Information*

We invite you now to answer some questions for your learning assessment and to give us your feedback about the Learning Object by evaluating its quality with Evaluate this Learning Object: Digitisation Lyfe Cycle.

For further details:

Readings and training materials

Readings and training materials

« Back Forward »

Linked Heritage LEARNINGObjects

LINKING CULTURAL HERITAGE INFORMATION

5/5 – CREDITS

Project by:
Cristiana Bettella
Elena Bianchi
Lisetta Dainese
Massimo Ferrante
Michele Visentin
Antonella Zane

CAB, University of Padua (Italy)

Texts by Michele Visentin
CAB, University of Padua (Italy)

Graphic and Web design by Gianluca Drago
CAB, University of Padua (Italy)

Funded by the Competitiveness and Innovation Framework Programme, 2011.

« Back

Linked Heritage
LEARNINGObjects

GENERAL OVERVIEW

The Linked Heritage trainin[illegible]me in Moodle

The main site which, besides the links to learning objects listed here, includes further readings and training materials, alongside with assessment and evaluation questionnaires.

» The Linked Heritage training programme in Moodle

The training programme includes four modules (syllabi) specifically tailored for the following target user groups:

- Cultural institutions decision makers
- Teachers and researchers
- LIS professionals
- Market players

Learning objects

Learning objects considered individually.

1. Persistent identifiers: What if?

It deals with Persistent identifiers (PIDs). What would happen if PID functional requirements were not met? Follow this audiovisual metaphor.

- Learning object
  - English
  - French
  - German
  - Greek
  - Polish
  - Spanish
- Linked Heritage training programme in Moodle

2. Digitisation life cycle

It presents the digitisation workflow both theoretically and practically. The first part gives an overview of the entire digitisation workflow, while the second one focuses on a case study of the University of Padova. Two videos are included.

- Learning object
  - English
  - German
  - Greek
  - Italian
  - Polish
  - Spanish
- Linked Heritage training programme in Moodle

3. MINT services

Metadata mapping procedures through MINT, the technological platform developed by the National Technical University of Athens.

It also gives practical tips to technicians working on mapping activities.

- Learning object
  - English
  - Spanish
  - Swedish
- Linked Heritage training programme in Moodle

4. Why and how to contribute to Europeana

It describes why cultural heritage institutions should contribute to Europeana, and also the University of Padova experience.

- Learning object
  - English
- Linked Heritage training programme in Moodle

5. Persistent Identifiers: Commercial and heritage views

A set of short case studies from both the commercial media sector (book publishing and retail) and the cultural heritage sector (libraries and research data archives). Learning Objects enable library and information professionals, as well as museum, archive and other heritage professionals to engage with the best practice in the commercial world and find common grounds, in view of potential future public-private partnerships.

- Learning object
  - English
  - Greek
- Linked Heritage training programme in Moodle

6. Public-Private Partnership with Europeana

A very brief overview of the "state of play" for commercial companies wishing to offer their product data to Europeana, with pointers to freely-available tools and documents contributed by EDItEUR to support this process as part of the Linked Heritage.

- Learning object
  - Bulgarian
  - English
  - Spanish
  - Swedish
- Linked Heritage training programme in Moodle

7. Terminology

This LO has been developed in the framework of WP7 of the Linked Heritage project.
Its aim is to introduce work package 3 of the Linked Heritage-project to unskilled or entry level learners in galleries, libraries, archives and museums.
Work package 3 is dedicated to the exploitation of monolingual or multilingual vocabularies in a digital environment.

- Learning object
  - English
- Linked Heritage training programme in Moodle

8. Linking Cultural Heritage Information

This LO focuses on the current structure of the Linked data Cloud - the best known representation of linked data. Size, licenses, subjects... all the informations were extracted from the analysis made by *Linking cultural heritage information*, the 2nd Work package of *Linked Heritage*. At the end, Cultural heritage institutions dealing with linked data publication (museums, libraries, archives..) can find best practice recommendations.

- Learning object
  - English
- Linked Heritage training programme in Moodle


cip
LINKED HERITAGE
europeana
think culture
UNIVERSITÀ DEGLI STUDI DI PADOVA

YouTube IT

Upload | LinkedHeritageTrain | 0

LinkedHeritageTrain

DASHBOARD

VIDEO MANAGER
- Uploads
- Playlists
- Tags
- Copyright Notices
- Search History
- Favourites
- Likes

CHANNEL SETTINGS

ANALYTICS

INBOX

Send feedback

Uploads 45

Search uploads

Actions | Playlists | Tag | View: Newest

	MINT Services - 5/5 HD 12 June 2013 12:39 Edit · Improve this video		4 0 0 0
	MINT Services - 4/5 HD 12 June 2013 12:36 Edit · Add music		5 0 0 0
	MINT Services - Rights 3/5 HD 12 June 2013 12:29 Edit · Add music		10 0 0 0
	MINT Services - Rights 2/5 HD 11 June 2013 16:46 Edit · Add music		7 0 0 0
	MINT Services - Rights 1/5 HD 11 June 2013 16:05 Edit · Add music		20 0 0 0
	Linked Heritage Training - Europ… 20 May 2013 16:35 Edit		30 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:25 Edit		37 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:24 Edit		45 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:24 Edit		43 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:23 Edit		40 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:22 Edit		42 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:21 Edit		40 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:21 Edit		40 0 0 0
	Persistent Identifiers: What if? » 1… 5 February 2013 11:20 Edit		45 0 0 0

	Persistent Identifiers: What if? » 9···	5 February 2013 11:19	Edit	47 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 8···	5 February 2013 11:19	Edit	47 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 7···	5 February 2013 11:18	Edit	48 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 6···	5 February 2013 11:17	Edit	51 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 5···	5 February 2013 11:16	Edit	51 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 4···	5 February 2013 11:16	Edit	58 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 3···	5 February 2013 11:15	Edit	62 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 2···	5 February 2013 11:13	Edit	90 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 1···	5 February 2013 11:10	Edit	104 views, 0 likes, 0 dislikes, 0 comments
	Europeana stopmotion	29 January 2013 15:36	Edit	37 views, 0 likes, 0 dislikes, 0 comments
	Linked Heritage Training -- Prom···	29 November 2012 14:58	Edit	6 views, 0 likes, 0 dislikes, 0 comments
	Linked Heritage Training -- Prom···	28 November 2012 12:21	Edit	30 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 1···	7 November 2012 14:51	Edit	17 views, 0 likes, 0 dislikes, 0 comments
	Persistent Identifiers: What if? » 8···	31 October 2012 11:48	Edit	32 views, 0 likes, 0 dislikes, 0 comments