



The Digital Cultural Heritage use case

Antonella Fresa, Promoter Srl
Rosette Vandenbroucke, BELSPO

Table of Content



- DCH use case
- DP Roadmap for DCH
- Sustainability
- Education and training needs for the DCH sector

Background



- ❑ The European amount of **digitized material** is **growing** very rapidly, as National, regional and European programmes support the digitization processes by Museums, Libraries, Archives, Archaeological sites and Audiovisual repositories
- ❑ Each country, each sector and often each organisation have **different policies, procedures and guidelines** for accessing, sharing, managing and preserving the content
- ❑ Recent studies commissioned by the EC showed that:
 - 83% of institutions said curatorial care is part of the mission
 - 83% of institutions have a digital collection or is currently involved in digitisation activities
 - 20% of all collections, that need to be, are digitised
 - 89% of audio visual institutions have born digital collections, while 43% of museums of art and history have them
 - 34% of institutions have a written digitisation strategy
 - About one third of the institutions are included in a national digitisation strategy, for national libraries more than half are included

Sources

- NUMERIC Study Report: http://cordis.europa.eu/fp7/ict/telearn-digicult/numeric-study_en.pdf
- ENUMERATE Survey Report on Digitisation in European CH Institutions 2012: <http://www.enumerate.eu/fileadmin/ENUMERATE/documents/ENUMERATE-Digitisation-Survey-2012.pdf>
- EC Comité des Sages Report on Cost of Digitising Europe's CH: http://ec.europa.eu/information_society/activities/digital_libraries/doc/refgroup/annexes/digiti_report.pdf

DCH needs



- ❑ High quality information technology management, to ensure **trust, availability, reliability, long-term safety of content, security, preservation and sustainability**
- ❑ Enhanced **access** facilities
 - to the researchers who will look for contents into the DCH e-Infrastructure for their research
 - to the cultural institutions that will deliver their data to the DCH e-Infrastructure
- ❑ **Interoperation** among existing cultural heritage repositories, among cultural portals and among data from the digital cultural heritage and from the research

Main challenges



- ❑ High investment for the production of DCH data due to the need of **human intervention** of experts
- ❑ High costs of digital preservation, due to the use of **separate solutions** implemented by each memory institution
 - The estimated total cost of digitising the collections of Europe's museums, archives and libraries, including the audiovisual material they hold is approximately €100bn, or €10bn per annum for the next 10 years
 - The cost of preserving and providing access to this material over a 10-year period after digitisation would be in the order of €10bn to €25 bn, provided that "federated" repository infrastructure is made available for the purpose
- ❑ DCH content is **complex** and **interlinked** through many relations
- ❑ **Contextual** data are very important for cultural research
- ❑ The digitisation process is **unique** cannot be replicated unless the whole work is done from scratch

e-infrastructures and DCH



- ❑ 2 twin-projects (**DC-NET** and **INDICATE**)
- ❑ an ongoing international coordination action (**DCH-RP**) brought together in the last years memory institutions and e-infrastructure providers from all over Europe to work for the future, in order to create a data infrastructure devoted to cultural heritage research
- ❑ These initiatives are contributing to shape the way towards an **Open Science Infrastructure for Digital Cultural Heritage**

Benefits offered by the e-infrastructures to DCH



- ❑ To allow for **cost reduction** in digitisation, cataloguing and metadata generation by substituting expensive human workforce with cheaper machine processes
- ❑ To support the **permanent identification** of digital cultural objects and providers
- ❑ To facilitate **storage and preservation**, ranging from short-medium- to long-term
- ❑ To improve **search facilities** supporting semantic search and linked open data
- ❑ To enhance **processing and visualisation of complex cultural data** (e.g. 3D modelling and VR representations) through the computing resources offered by research e-infrastructures
- ❑ To enable dynamic distributed **virtual organisations**, facilitating collaboration with information and resource sharing (e.g. virtual conferences, document sharing, blog, cooperation platforms, ...)
- ❑ To contribute to **standardisation** in the data world, e.g. by developing a common reference model for the DCH sector

Other EC co-funded initiatives



❑ **Europeana Photography**

- CIP ICT PSP Pilot B action putting together and sending to Europeana some of the most prestigious photographic archives, public libraries and photographic museums covering the length of time from the beginning of photography to the beginning of the Second World War (1839-1939)
- Special attention is devoted to the management of intellectual property, which is further emphasised by the involvement of content providers from both the private and the public sectors

❑ **PREFORMA**

- FP7 Pre-Commercial Procurement to address the challenge of implementing good quality standardised file formats for preserving data content in the long term
- The main objective is to give memory institutions full control of the process of the conformity tests of files to be ingested into archives

❑ **Europeana Space**

- Best Practice Network with the aim to increase and enhance the creative industries' use of DCH content by delivering a range of resources to support their engagement
- Europeana Space addresses all sectors of the creative industries, from content providers to producers, exhibitors, artists and makers of cultural/creative content, publishers, broadcasters, telecoms and distributors of digital content
- The project aims to address the problems, which limit the re-use of Europeana by the creative industries, such as issues around the IPR status of content and the need for business models demonstrating the potential for exploitation of available content

Data types and formats



- ❑ DCH content is composed of **several different typologies** of information and **different formats**: texts, still images, 3D models, publications, digital exhibitions, virtual reconstructions, etc.
- ❑ Examples of standardised formats often used by memory institutions are:
 - **Documents.** Public authorities and other institutions producing electronic documents and media content on national level are normally using open standards adapted to specific requirements to produce their electronic files. **PDF/A**, and its different versions, is for example the standard mostly used by archiving institutions for electronic documents
 - **Images.** **TIFF** and **JPEG2000** are the preservation format most often used by memory institutions for still image digitisation
 - **Audiovisual** contents. The Material eXchange Format (**MXF**), the container format (developed and maintained by audio-visual industry, particular for postproduction and distribution purposes) + **M-JPEG200** and **FFV1** for the actual coding

Data retention and storage



- ❑ In the DCH sector, data which are digitised are then retained. The main **retention requirements** are:
 - Separation of content and metadata
 - OASIS compliance
 - Accessibility through retrieval and search system
- ❑ Generally, data and metadata are stored in **data centres / repositories hosted by the memory institutions themselves**
 - this poses big maintenance issues due to the lack of ICT expertise
- ❑ Access to shared resources is very appealing, but there are still problems in adopting this approach:
 - Issues related to **copyrights**
 - Cultural data are curated by many different persons: **data management and administration + user access control** are very important
 - **Security** of the data is very important for cultural institutions: trust building is a key factor when it is not determined where data are stored
 - Functionalities and services offered by e-infrastructures should not impact on the **outgoing traffic** of the institution
 - Access to the e-infrastructure services should be simple without requiring **IT specialist knowledge**

Use of standards



- ❑ The extensive use of **relevant and open standards** is a vital pre-requisite for the CH community to promote interoperability, encourage widespread access and control costs in its digital preservation programmes.
- ❑ Extensive reviews under the auspices of the Minerva (2008), Athena (2009), Linked Heritage (2011) and DCH-RP (2013) projects categorized and described many of the standards that are most applicable or recommended in this area.
 - Examples are: EAD, OAIS, ONIX, Indecs, EDM, Premis, CIDOC-CRM, FRBR, EPIDOC, Dublic Core, SKOS; MARC, METS, MAB2, MODS, Museumdat, ObjectID, SPECTRUM, LIDO, BIBO, etc.
 - Furthermore, the OAI-PHM/OAI-DC standards are used to aggregate data and to make data available for publication in other portals (as in the case of Europeana).

Persistent identification



- ❑ The **PID requirements** do not vary significantly from one DCH initiative to another – this represents a service useful to most DCH work, ‘out of the box’.
- ❑ This topic was covered at some length within Deliverable D2.2 of the Linked Heritage project. Primary candidate identifier types for use in the Cultural Heritage arena are:
 - general digital identifier standards: **URI** (Universal Resource Identifier), **URL** (Universal Resource Locator) and **URN** (Universal Resource Name)
 - service-associated digital identifier standards: **PURL** ((Persistent URL) & Handle System), **DOI** (Digital Object Identifier), **OpenURL** and **ARK** (Archival Resource Key)
- ❑ Arguably, the service-associated and maintained identifiers are likely to offer more comprehensive features to CH institutions managing digitized resources, but issues relating to both **cost** and **policy** have militated against the widespread adoption of such identifiers in this area.

Long term preservation



- ❑ Long-term preservation of digital cultural content has been identified as the **highest priority** for the DCH sector (crf. Service Priority Handbook produced by DC-NET)
 - 23% of institutions have a written digital preservation strategy, figures range from 44% for national libraries to 12-25% for museums
 - About a third of the institutions are included in a national preservation strategy
 - 40% of national libraries say there is no national digital preservation strategy
 - 30% of the institutions are included in a national digital preservation infrastructure
- ❑ Type and size of content to be preserved vary from case to case
 - **Types** include: texts, still images, 3D models, publications, digital exhibitions, virtual reconstructions, etc.
 - **Size** range from 5 to 200 GB

Access to data



- ❑ The creation of the “location” for the **online presentation** of DCH materials (a content management system, a portal system, a digital library or digital repository) is a central part of any digital heritage initiative
- ❑ **Authentication and Authorisation** requirements:
 - Generally, most of the data are available to others, apart from those that are not totally documented
 - Usually, access to these data is open for view only; protected instead for importing and updating data
 - Adding and editing data needs to be password protected and limited to known individuals authorised by the institution
 - Authentication mechanisms most in use: Open access, Password protected, IP-based, X.509 certificate based, Shibboleth or equivalent
- ❑ **Federated access** can be a valid approach to reduce the number of credentials for the users, increase security and improve the users experience offering single sign on: sign in once, access more resources

DP Roadmap for DCH

The starting point



Current solutions normally require:

- adaptation to the specific mandate of the individual cultural heritage institution,
- its existing technological infrastructure,
- the competences of its staff.

Twofold scope of the DCH-RP Roadmap:

- to help policy makers and programme owners to plan ahead
- to assist managerial teams of cultural heritage institutions in taking decisions related to digital preservation

Initial steps towards a Roadmap



Unlike digitisation, where common approaches and best practices are rather well developed, digital preservation is still an area where workflows and easily applicable universal toolkits are not widely available

A range of initial steps are considered necessary to prepare the actual implementation of a DP Roadmap for DCH:

I. Glossary

- In order to share common areas of development between communities, it is important to establish a common taxonomy

II. Registry of preservation services that are already available and how they can be exploited

III. Standards

- data formats, metadata schema, skos-compliant multilingual vocabularies, aggregation protocols

IV. Trust model for the use of e-Infrastructures by DCH users

A DP Roadmap for DCH



To develop a commonly agreed **vision** of distributed digital preservation architecture relying on e-Infrastructures

Four areas identify the policy domains that require intervention:

- ❑ **Harmonisation** of data storage and preservation
- ❑ Improved **interoperability**, including integration of preservation within the overall workflows for digitisation and online access
- ❑ Establishment of conditions for **cross-sector integration**, transferring knowledge and know-how
- ❑ **Governance** models for infrastructure integration, to foster participation in larger e-Infrastructure initiatives, and aggregation and re-use of digital resources

The timeframe



Important: **flexibility**, to make it possible for each cultural heritage institution to define its own practical action plan.

To develop a matrix, where for each element of our vision, we map the activities along a timeline:

- short-term action plan (now)
- Medium-term (2016)
- Long-term (2020)

An infrastructure model for distributed DP



- ❑ **Data infrastructure** (see EUDAT services)
- ❑ **Service architecture**
 - Functional areas, following the OAIS model and the actual current preservation process (pre-ingest, ingest, archival storage, preservation planning, data management, access and use) to serve specifically the needs of the DCH community
 - Types of services and objects addressed
 - Level of maturity
 - Licensing conditions

Sustainability

Data re-use



- ❑ The end users of DCH work tend to be **DCH researchers** and/or members of the **general public**
- ❑ Other use cases include re-use of data in **education**, in **commercial** ventures (e.g. publishing, tourism), in **collaborative projects** and for **digital exhibitions**
- ❑ In the last years, **Cultural and Creative Industries** revealed to be an interesting possibility to exploit the potentials of digital cultural content as an essential driver of creativity, innovation and competitiveness in the framework of a sustainable economy, as highlighted first by UNESCO and then by the European Commission
 - **Cultural Industries** comprise the economic domains that focus on Cultural Heritage, museums and libraries, cultural and heritage tourism and also education as well as research in cultural domains
 - **Creative Industries** comprise domains such as arts (visual and performing arts) and architecture, design, crafts, fashion, music, film, publishing, advertising, TV and Radio, toys, video games and serious-educational games, software, as well as research and development in technological domains (R&D)

Business models



- ❑ Memory Institutions need to digitize their content primarily for **preserving** it in a digital format and for granting and enlarging the **access** to them by researchers, teachers & students and citizens. These are public services that need to access to public funding for their sustainability.
- ❑ Cultural institutions should become both **content providers** and **service providers**, exploring new audiences and markets and promoting further investment in digitisation of cultural content
- ❑ New projects have been funded in the last years by the EC (such as Europeana Photography and Europeana Space) to experiment with innovative applications and services the **creative re-use** of cultural resources

Involving cultural and creative industries



- ❑ The use of cultural heritage content by the creative industries is still limited by factors including:
 - issues around the **IPR** status of content
 - poor metadata **quality**
 - the need for **business cases** demonstrating the potential for exploitation of digital cultural content
 - **awareness** and education of the cultural heritage sector about the exploitability of the cultural assets in the digital world

Achieving impact



- Some basic requirements to be served:
 - To define clear and simple **rules** on how to **use and re-use** data
 - To create a **collaboration environment** where memory institutions, creative industries and end users can meet and interact with e-Infrastructure providers, such as www.digitalmeetsculture.net reaching c 15,000 visitors per month
 - To develop spaces of **business and innovation** where products and services can be promoter, also through **public-private partnerships**
- Target users to be involved:
 - The **education sector**, with new and innovative e-learning products based on the available digital resources
 - The **humanities research sector**, with new opportunities to develop new advanced services and applications based on innovative technologies
 - The **citizens**, who can be stimulated by new interactive ways to access and engage with the cultural heritage (e.g. cultural tourims, citizen scientists)

Educational and training needs in the DCH sector



- ❑ ICT, metadata architectures, data modelling, DB management, web engineering, new media management, data warehouses and repositories, W3C/ISO standards and technologies, interoperability/access protocols, data mining, filtering
- ❑ Digitization techniques
- ❑ Intellectual Property Management
- ❑ Storage, archiving, preserving methods and strategies
- ❑ Information design, web design, usability engineering, user studies, web services, CRM, communication strategies
- ❑ Digital library management, collection management, collaborative/federated approaches, content management, project mgmt., cross-cultural and multilingual issues
- ❑ Knowledge organization, ontologies, advanced indexing and retrieval methods,
- ❑ Visualization, 3D modelling, augmented reality, gamification.

Next Appointments



❑ 4 April 2014, Brussels

- Information Day to present the new procurement launched by PREFORMA to check conformance of files against standards (<http://www.preforma-project.eu/info-day.html>)

❑ 23-24 April 2014, Tallin

- DCH-RP Concertation meeting between cultural heritage institutions and e-Infrastructure providers (<http://www.digitalmeetsculture.net/article/dch-rp-e-infrastructure-concertation-workshop/>)

❑ 20 May 2014, Helsinki

- Workshop @ EGI Community Forum 2014: E-infrastructures and services for data preservation and curation (<http://www.digitalmeetsculture.net/article/e-infrastructures-for-data-preservation-and-curation-egi-cf-2014-helsinki/>)



Thank you!

Antonella Fresa
Promoter Srl
[*fresa@promoter.it*](mailto:fresa@promoter.it)
[*www.digitalmeetsculture.net*](http://www.digitalmeetsculture.net)

Rosette Vandenbroucke
BELSPO
[*rosette.vandenbroucke@gmail.com*](mailto:rosette.vandenbroucke@gmail.com)