Feedback form for user communities: the Digital Cultural Heritage (DCH) case

# Scientific community

We are providing information related to the Digital Cultural Heritage (DCH) sector.

The volume of digital cultural heritage data is incredibly growing year after year, so now it is necessary to reflect upon the tools which permit to manage such a huge amount of data in an efficient and selective way, in order to make the data available to the researchers and the citizens in a European dimension.

The needs of DCH sector are:

* 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 are:

* High investment for in 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 centralised repository infrastructure is made available for the purpose
* DCH content is difficult to be preserved because data are 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.

2 twin-projects (DC-NET and INDICATE) and 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. In particular, long-term preservation of digital cultural content has been identified as the highest priority for the DCH sector.

The potential benefits from the use of e-infrastructure are:

* 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- and long-term
* To improve search facilities to manage 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 (grid, cloud)
* To enable dynamic distributed virtual organisations, facilitating collaboration with information and resource sharing (e.g. virtual conferences, document sharing, blog and cooperation platforms, etc.)
* To contribute to standardisation in the data world, e.g. by developing a common reference model for the DCH sector

These initiatives are contributing to smooth the way to the Open Science Infrastructure for Digital Cultural Heritage which is foreseen in 2020.

Other valuable initiatives in this field are:

* PREFORMA (PREservation FORMAts for culture information/e-archives) a Pre-Commercial Procurement (PCP) co-funded by the European Commission under its FP7-ICT Programme 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 Photography, a Pilot B action putting together and sending to Europeana some of the most prestigious photographic archives, public libraries and photographic museums covering specifically the length of time from the beginning of photography (1839 with the first example of images from Fox Talbot and Daguerre) to the beginning of the Second World War (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.
* Europeana Space, a Best Practice Network with the aim to increase and enhance the creative industries’ use of Europeana by delivering a range of resources to support their engagement. Europeana Space will address 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.

# Contacts

Antonella Fresa, Promoter Srl, [fresa@promoter.it](mailto:fresa@promoter.it)

Claudio Prandoni, Promoter Srl, [prandoni@promoter.it](mailto:prandoni@promoter.it)

# Data management planning

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.

Recent studies[[1]](#footnote-1) 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

CH institutions often do not have a well defined data management plan.

Each country, each sector and often each organisation have different policies and guidelines for accessing, sharing and processing the content under its control.

DCH content is composed of several different kind 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:

* Document formats. 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.
* Image formats. TIFF is the preservation format most often used by memory institutions for still image digitisation.
* Audiovisual formats. The Material eXchange Format (MXF) is a container format for professional digital video and audio media which is developed and maintained by audio-visual industry, particular for postproduction and distribution purposes + M-JPEG200 and FFV1for the actual encoding.

# Persistent ID bridge

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’. There is excellent research and development already done in this area, and one or more of the existing schemes (e.g. DOI, ARK, URI, URN, etc.) could be adopted with minimal adjustment.

This topic was covered at some length within Deliverable D2.2 of the Linked Heritage project. The report explained the role and importance of such identifiers before describing the primary candidate identifier types for use in the Cultural Heritage arena. Fundamental features of three general digital identifier standards – URI (Universal Resource Identifier), URL (Universal Resource Locator) and URN (Universal Resource Name) – were presented first and they are included in the table in section 4.1.5 above. The report then described four types of service-associated digital identifier standards: PURL ((Persistent URL) & Handle System), DOI (Digital Object Identifier), OpenURL and ARK (Archival Resource Key). Summary information on these four is presented in the table below.

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.

The example of DOIs may be illustrative here. Although the utility of this identifier system is clear, uptake to date in the DCH environment has been very low. The costs associated with DOI registration may represent one significant barrier. But equally, the absence of a DOI Registration Agency specifically aligned with the DCH community – understanding cultural heritage requirements and able to design registration metadata that is meaningful to that community – may represent an even greater impediment. (Compare for instance the near-ubiquity of DOIs in the area of scientific, technical and medical journals, which are serviced by the Crossref registration agency, itself originally established by the STM publishing industry itself.)

# Metadata packaging

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. The more relevant deliverables from the earlier projects are available as follows:

* Athena: <http://www.athenaeurope.org/index.php?en/149/athena-deliverables-and-documents>
  + *D3.1, Report on Existing Standards Applied by European Museums.*
  + *D3.2, Recommendations and Best Practice Report.*
* Linked Heritage: <http://www.linkedheritage.eu/index.php?en/142/documents-and-deliverables>
  + *D2.1, Best practice report on cultural heritage linked data and metadata standards.*
  + *D2.2, State of the art report on persistent identifier standards and management tools.*
* DCH-RP: <http://www.dch-rp.eu/index.php?en/61/deliverables>
  + *D3.2, Standards and interoperability best practice report.*

Examples of some of the most relevant and used standards in the CH sector are: EAD (Encoded Archival Description), ISAD (General International Standard Archival Description), OAIS (Open Archival Information System), MoReq2 (Model Requirements Specification for the Management of Electronic Records), ISAAR (International Standard Archival Authority Record for Corporate Bodies, Persons and Families), ONIX for Books, VCARD, Indecs (Interoperability of data in e-commerce systems), DBpedia Ontology, EDM (Europeana Data Model), VRA Core (Visual Resources Association), Premis (Preservation Metadata: Implementation Strategies), CIDOC-CRM (CIDOC Conceptual Reference Model), Dublin Core, SKOS (Simple Knowledge Organization System), Basic Geo, MIDAS Heritage, FRBR (Functional Requirements for Bibliographic Records), MAB2 (Maschinelles Austauschformat für Bibliotheken), MARC (MAchine Readable Cataloging), METS (Metadata Encoding and Transmission Standard), MODS (Metadata Object Description Schema), CDWA (Categories for the Description of Works of Art), museumdat, Object ID, SPECTRUM, LIDO (Lightweight Information Describing Objects), MO (Musical Ontology), MusicXML, FOAF (Friend Of A Friend), BIBO (Bibliographic Ontology).

Furthermore, the OAI-PHM/OAI-DC standards are used in certain cases to aggregate data and to make data available for publication in other portals (as in the case of Europeana).

# Data access support

Usually, data and metadata are stored in data centres / repositories hosted by the memory institutions themselves, even if this poses big maintenance issues due to the lack of ICT expertise.

Studies conducted in the projects mentioned above showed that access to shared resources, as in the case of the e-infrastructures (cloud, grid), is very appealing. The main problems in adopting this approach are:

* 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.

The creation of an online “location” for the presentation of DCH materials online is a central part of any digital heritage initiative. The model applied will most commonly be a content management system, a portal system, a digital library or digital repository which has been specifically designed and built for the purpose.

Access requirements:

* Generally, most of the data should be available to others
* 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, Shibboleth or equivalent

Federated access can be a valid approach:

* To reduce the number of credentials for the users,
* To increase security
* To improve users experience (sign in once, access more resources)

# Data re-use and value proposition

Concerning re-use

* 31% of the institutions have a policy on the use of the digital collections, figures range from 60% for national libraries to 22% for archaeology museums
* 42% of institutions monitor the use of their digital collection
* 85% of institutions use web statistics to measure the use of their digital collections
* By 2014 institutions estimate to make twice as much of their collections accessible through Europeana when compared to today

‘Users’ of digital cultural heritage data are researchers, curators and the general public

‘Re-users’ are Cultural and Creative Industries, which revealed to be in the last years 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 museums, libraries, cultural tourism, as well as education and research in cultural domains
* Creative Industries comprise arts (visual and performing arts), architecture, design, crafts, fashion, music, film, publishing, advertising, TV and Radio, toys, video games and serious-educational games.

Use cases for the re-use of DCH exists for: educational products, commercial ventures (e.g. publishing, tourism), collaborative social projects and digital exhibitions.

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
* Successful business cases demonstrating the potential for exploitation of digital cultural content
* Lack of awareness by the cultural heritage sector about the exploitability of the cultural assets in the digital world

# Value proposition and business planning

Memory Institutions need to digitize their content primarily for preserving it in a digital format and for granting and enlarging access to them by researchers, teachers & students and citizens. These are public services that need public funding for their sustainability.

The problem is that memory institutions’ budgets often don’t meet the requirements of such projects. In search of funding issue solution cultural heritage institutions tried many variants – lobbying activities to establish additional governmental funding, participation in the projects of international organizations, partnership with other memory institutions to share the costs of digitization, seeking sponsor help and so on.

Museums, libraries, archives should become content providers and service providers, exploring new audiences and markets and attracting further investment in digitisation of cultural content. New projects have been funded in the last years by the EC (such as EuropeanaPhotography and EuropeanaSpace) to experiment with innovative applications and services for the creative re-use of cultural resources.

In this context, public-private partnerships have also been taken into account as possible ways to exploit the digital resources made available by the memory institutions, as demonstrated by the establishment of dedicated task forces in Europeana and in other related projects (such as Linked Heritage and Europeana Photography).

The embracing of the e-Infrastructures by the digital cultural heritage community can open new scenarios of use and exploitation with an impact expected on different sectors:

* The cultural heritage sector. The managers who work in the Cultural sector can become more aware about the potential that the e-infrastructures can offer to their work: storage, preservation, access services for the cultural institutions, etc.
* The research. A better integration of the cultural sector with the e-Infrastructures can enable the research of new advanced services and applications.
* The economic sectors. Digital cultural content can become more usable and re-usable for education, cultural tourism, long-life learning, non-professional cultural interests, creative industry, etc.

The most important measures to be taken to achieve the expected impact are:

* To provide information, news, links to knowledge resources that can support the dialogue between different actors
* To understand the changes generated by digitisation in the society
* To engage citizens in digital cultural heritage
* 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.netreaching c 15,000 visitors per month
* To develop spaces of business and innovation where products and services can be promoted, also through public-private partnerships

The target users to be involved are:

* 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 tourism, citizen scientists)

# Educational and Training needs

The main educational and training needs for the DCH community can be summarised as follows:

* Use of the e-infrastructures
* 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
* High technology illustrations such as 3D modelling, augmented reality, etc.

1. 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> [↑](#footnote-ref-1)