| | |
|---|---|
| **Project Acronym:** | INDICATE |
| **Project Title:** | International Network for a Digital Cultural Heritage e-Infrastructure |
| **Contract Number:** | 261324 |
| **Starting date:** | 1 September 2010 **Ending date:** 31 October 2012 |

| | |
|---|---|
| **Deliverable Number:** | D 3.6 |
| **Title of the Deliverable:** | Best practice for applying research pilots and use case studies to digital cultural heritage |
| **Task/WP related to the Deliverable:** | WP3/ Task 3.3 Guidelines, standards and best practice |
| **Type (Internal or Restricted or Public):** | Public |
| **Author(s):** | all the partners |

| | |
|---|---|
| **Contractual Date of Delivery to the CEC:** | 26 |
| **Actual Date of Delivery to the CEC:** | 5 November 2012 |

**Project Co-ordinator**


| | |
|---|---|
| *Company name :* | Istituto Centrale per il Catalogo Unico (ICCU) |
| *Name of representative :* | Rosa Caffo |
| *Address :* | Viale Castro Pretorio 105, I-00185 Roma |
| *Phone number :* | +39.06.49210427 |
| *Fax number :* | +39.06.4959302 |
| *E-mail :* | rcaffo@beniculturali.it |
| *Project WEB site address :* | http://www.indicate-project.eu |

## Context

| | |
|---|---|
| WP 3 | Dissemination |
| WP Leader | CULTNAT |
| Task 3.3 | Best practice for applying research pilots and use case studies to digital cultural heritage |
| Task Leaders | ICCU (T3.3) |
| Dependencies | WP2 (Network of Common Interest), WP4 (Research Pilots), WP5 (Research Case Studies) |
| Starting date | Month 1 |
| Release date | Month 26 |

| | |
|---|---|
| Reviewers | NTUA, GARR, COMETA |

**HISTORY**

| Version | Date | Author | Comments |
|---|---|---|---|
| 0 | 30/5/2012 | Fresa | First draft sent to the mailing list for comments and further discussed at the project meeting in Paris on 8/6/2012 |
| 1 | 18/6/2012 | Fresa | Second draft sent to the mailing list for comments |
| 2 | 23/8/2012 | Fresa | Third draft including the comments received from NTUA and COMETA |
| 3 | 23/8/2012 | Fresa | Final version including the comments received from GARR |
| 1.0 | 5/11/2012 | Fresa | Final version submitted to EC |

# Table of Contents

# 1 Executive Summary

This deliverable, organised as a handbook, provides information about pilot projects and case studies in digital cultural heritage across Europe, as assembled by the INDICATE project. The handbook includes 6 main chapters, in addition to this Executive Summary. Chapter 2 gives an **introduction** to e-infrastructure and digital cultural heritage (DCH), as well as to research pilots and use case studies. Chapter 3 gives a **background** overview of e-infrastructures and digital cultural heritage, and background related to research pilots and research use case studies. Chapter 4 describes the approaches, results and best practice for the INDICATE **research pilots** for semantic searching and e-collaborative digital archives. Chapter 5 details the results and best practice which arose from the INDICATE **research use case studies** for long-term digital preservation, virtual exhibitions and performances, and geo-coded digital cultural content. Chapter 6 describes best practice for the application of **research pilots** and use case studies to DCH. Chapter 7 gives the **conclusions** of the handbook, specifically focusing on the use of e-infrastructures by the DCH community.

# 2 Introduction

This present handbook summarises the results and best practice findings from the application of Research Pilots (WP4) and Research Case Studies (WP5) to e-Infrastructures for DCH. It is intended to demonstrate the usefulness of these research methods for validating theoretical approaches and gaining new knowledge, and establish standards and best practice for future use of these methods. In addition, the handbook presents best practice for semantic searching, e-collaborative digital archives, long-term digital preservation, virtual exhibitions and geo-coded digital cultural content as identified by the research pilots and research use case studies carried out during the INDICATE project.

# 3 Background

E-Infrastructures offer many benefits to the DCH community, including stable technical platforms, federated authentication and authorisation infrastructure, cost-effective data storage and superior processing power. While the use of e-Infrastructures in the DCH sector has increased in recent years, **many DCH organisations remain unaware of the services provided by existing e-infrastructures** at the national and international level and the potential for new services which leverage these e-Infrastructures. It is important that the DCH sector effectively exploits the potential of e-infrastructures in order to maximise efficient use of resources in the future and to fully benefit from the advantages provided by e-infrastructure based approaches.

During the INDICATE project, practical experimentation using the policies and the best practices identified by DC-NET,[1] INDICATE's sister project, was carried out using research pilots and use case studies. Research pilot studies are an effective way to test proposed applications of e-infrastructures before full-scale implementations, while research use case studies are useful for identifying potential future applications of e-infrastructures for DCH. This handbook details best practice for application of these research methods to DCH and e-infrastructures as identified by the INDICATE project. In addition, this handbook summarises the best practice identified using these methods for: semantic searching, e-collaborative digital archives, long-term digital preservation, virtual exhibitions and geo-coded digital cultural content.

[1] www.dc-net.eu

## 3.1 Research Pilots

Research pilot studies allow researchers to conduct preliminary tests of a process before carrying out full-scale studies. Pilot studies are used extensively in physical sciences and humanities research. Testing performed during pilot studies can be a scaled-down version of the entire process under examination, or can be a series of simple tests of selected parts of larger, more complicated processes. Carrying out pilot studies yields numerous benefits, including:

- **Reduced risk:** potential problems are identified early in the research process so they can be addressed before further work is carried out, thereby increasing the likelihood of success of full-scale studies/implementation;
- **Reduced cost:** performing small-scale studies costs less than full-scale studies/implementation;
- **Proof of concept:** the results of pilot studies can be used as proof-of-concept to secure approval or funding for full-scale studies/implementation;
- **Training:** pilot studies provide an excellent opportunity to train new researchers in a simpler, low risk environment than full-scale studies/implementation.

## 3.2 Research Use Case Studies

Research use case studies are used to empirically study a process, and are well suited to **generating and testing hypotheses**. The method is indeed used extensively in both the social sciences and life sciences. Use case studies involve in-depth analyses of specific, individual cases with the goal of improving understanding of a particular process or event. The use case study approach involves data collection, data analysis and result reporting. Data used for use case studies is often a **mix of quantitative and qualitative data**.

## 3.3 Sources of Content for the Handbook

The best practice identified for semantic searching, e-collaborative digital archives, long-term digital preservation, virtual exhibitions and geo-coded digital cultural content described here has been synthesised from a range of sources, including interactions with the Network of Common Interest (WP2) and the DC-NET project.[2] Content for the handbook has also been gathered from the following research pilots (WP4) and use case studies (WP5) carried out over the course of the INDICATE project:

- **Semantic searching pilot:** A pilot study of semantic search using e-Infrastructure was carried out using the MICHAEL dataset (D4.1). The pilot involved semantic transformation and enrichment of the dataset using cloud based infrastructure. The semantically enhanced dataset was also stored on the cloud platform in order to demonstrate the benefits of exploiting the processing power of e-infrastructures and the overall scalability of this approach. The pilot demonstrated that the scalability of the process was greater when deployed on the cloud infrastructure than on a single server, and that the cost was significantly lower;
- **E-Collaborative digital archives pilot:** A pilot study was carried out to demonstrate implementation of e-Collaborative Digital Archives on e-Infrastructures, illustrating mechanisms for protecting the archives with access control and rights management (D4.2). Three repositories were used to

[2] DC-NET is a "sister" project of INDICATE, focusing on DCH and e-Infrastructures across Europe. See dc-net.org

demonstrate the approach. The importance of standards, re-usability of components and ensuring ease of use for a range of users, including non-IT experts and "citizen scientists" were highlighted;

- **Long-term digital preservation case study**: A case study of long-term digital preservation was carried out to examine the current challenges faced in the area (D5.1). Best practices were identified, highlighting the need for global, standardized policies, effective staff training, persistent identifiers, long-term access to digital data and the potential benefits of employing e-infrastructures for long-term digital preservation;
- **Virtual exhibitions case study**: A case study was performed to determine the degree to which e-infrastructures are used for of virtual exhibitions and performances (D5.2). It was noted that virtual exhibitions and performances currently underutilize the capabilities of e-infrastructures. Best practices for virtual exhibitions and performances were identified and opportunities for future innovation were described;
- **Geo-coded digital cultural content case study**: A case study of geo-coded digital cultural content was carried out (D5.3). The potential of geocoding for enriching cultural material and increasing use of digital cultural items, especially in the area of cultural tourism, was highlighted. A practical examination of geo-parsing as a way to automatically add geographic coordinates to digital cultural items was also carried out, and demonstrated the effectiveness of the approach.

# 4 Research Pilots

## 4.1 Semantic Search Research Pilot

DCH collections are used to collect and manage digital objects. Initially, the focus of these collections has been the consumption of online content by human users. However, with advances in semantic search, the focus has shifted to computer-based knowledge retrieval and knowledge transfer. Therefore, dissemination and integration of large digital cultural collections must not ignore the semantics underlying the digital content. Approaches are also required the **enrichment of existing DCH collections to facilitate knowledge transfer through semantic searching**.

### 4.1.1 Approach

A pilot study was carried out using the MICHAEL dataset to investigate the suitability of cloud e-infrastructure for semantic search of digital cultural data. Briefly, the initial XML metadata were semantically enriched while being transformed to RDF in parallel. For the enrichment process, the values of specific elements of the dataset, such as country, person and language, were also discovered in external resources. The extra information added through the **enrichment process increased the searchability of the digital content**. The enriched RDF metadata was stored in a repository and used to support the semantic search of the MICHAEL content using information that was not available in the original dataset. The pilot workflow and repository were deployed on the Amazon Elastic Compute Cloud (EC2), using a processing interface (MINT-PI) that parallelized the semantic transformation and enrichment thereby leveraging the massive processing power of cloud infrastructure.

### 4.1.2 Results

The algorithm developed for locating external resources operated extremely well for the country and language parameters, while very good results were also obtained for the person parameter. A comparison of the overall performance of the workflow when deployed on e-infrastructures (i.e. using cloud system) versus on a single server showed that **distributed solutions can provide scalability at low cost** when processing massive volumes of content, such as for DCH repositories.

E-infrastructures can provide scalability at low cost when processing massive volumes of content, such as for DCH repositories.

### 4.1.3 Best Practices

#### 4.1.3.1 Exploit scalability and processing power of e-infrastructures (Grid and cloud computing)

Enrichment of metadata with information from external sources and semantic search of digital cultural repositories allows better use of digitised objects by a range of users, including DCH researchers and "citizen scientists". As demonstrated in the INDICATE pilot study, data from an existing DCH repository was enriched with information about people, countries and languages associated with digital cultural objects. This enrichment allowed users to search for digital objects associated with a specific person, country or language, improving the likelihood that the user would be able to locate relevant material in a straightforward manner.

While metadata enrichment and semantic search improve the searchability of material in DCH repositories, the processes are computationally expensive due to the large volume of content in the repositories and the requirement of the enrichment process to use network connections for the discovery of external resources. The total time, and hence the cost, of computations performed using e-infrastructures are a fraction of that required for single or locally clustered architectures, as demonstrated by the INDICATE semantic search pilot. Therefore, **e-infrastructures can be used to significantly improve the effectiveness and efficiency of computationally expensive processes**, such as enrichment of DCH content to enhance user interactions and experience.

E-infrastructures can be used to significantly improve the effectiveness and efficiency of computationally expensive processes.

## 4.2 E-Collaborative Digital Archives Research Pilot

E-Collaborative digital archives have a number of advantages over traditional archives. They are available from any location at any time, allow users access to fragile materials that would not otherwise be circulated, support long-term preservation efforts and can be organised to allow users to easily locate relevant content. While easy access to digital material is advantageous, measures should be taken to ensure protection of the digital cultural content and that the archives are accessible to both expert and non-expert IT users.

### 4.2.1 Approach

Three digital archives were used for the pilot study. The e-collaborative digital archives pilot had two aims:

- **Simplify access** for non-expert users;

- **Adoption of an authentication and authorization system, in accordance with standards and Research Community best practices,** to effectively protect digital cultural archive contents.

The pilot was carried out in the four following phases:

1) **Definition** of metadata used to describe each of the repositories, and ways that users should be able to access, browse and query the repository contents;
2) **Upload** of digital objects to e-infrastructure using several storage resources to achieve fault tolerance, balance loads in cases of high traffic, hide network latencies and allow users to choose a resource close to their location;
3) **Registration** of metadata for the uploaded digital objects in the three different repositories;
4) **Creation** of simple front-end access to the three repositories through the e-Culture Science Gateway (e-CSG)[3] using portlets for user access, browsing and retrieval.

### 4.2.2 Results

**Simplified user access** was achieved using a gLibrary platform used to create repositories on the storage resources and metadata service of a Grid infrastructure. gLibrary APIs were used to create a set of portlets, deployed on the e-CSG implemented with Liferay[4] to provide an easy-to-use front-end for discovering, finding and retrieving assets of the three digital archives used for the pilot. **User authentication and authorization** was achieved by integrating the Science Gateways single-sign-on mechanism with Identity Federations and user tracking and logging for any Grid transaction.

The e-CSG approach demonstrated its value in the simplification of user interfaces and of authentication. This encourages DCH uptake of e-Infrastructures

### 4.2.3 Best Practices

#### 4.2.3.1 Ease access to and use of e-infrastructures for non-expert users

Traditionally, e-infrastructures have been accessed using complex security mechanisms accessed through non-graphical interfaces, which are not easily used by non-expert users. Easier access to e-infrastructures by non-expert users (including "citizen scientists") through mechanisms such as the e-CSG and Identity Federations is a major step forward towards **increasing the uptake of Grid technology by the DCH community**. The Science Gateway model, supports Identity Federations and Social Networks, and can revolutionize the use of Grid infrastructures and widen their potential user base. To test the social impact of the e-CSG, a community page[5] has been created on Facebook (Google+, LinkedIn and other networks will follow soon) and the e-CSG has been integrated with the Social Network environment.

Once access has been gained, **non-expert users can benefit further from easy-to-use interfaces** with Grid infrastructure, such as the portlets of the e-CSG, which have maximum re-usability and have already been adopted by other projects (e.g., agINFRA, CHAIN, DECIDE, EarthServer, EUMEDGRID-Support, and GISELA).

[3] http://indicate-gw.consorzio-cometa.it
[4] www.liferay.com
[5] www.facebook.com/pages/INDICATEeCultureScienceGateway-Community/189043984560271

#### 4.2.3.2 Protect digital cultural content

In order to protect the digital cultural content contained in digital archives, access to the archive must be controlled: a user must be both authenticated and authorized. The e-CSG carries out authentication and authorization using the state of the art standards and federative approaches.

# 5 Research Use Case Studies

## 5.1 Long-Term Digital Preservation

Long-term digital preservation of cultural heritage for future generations is an important responsibility of the DCH community. **Digitisation is an essential element of long-term preservation**, especially for works that are difficult to preserve physically, such as ancient manuscripts. However, care must be taken to ensure that the digitised versions of the cultural content are also preserved by maintaining the integrity and future accessibility of digital records.

### 5.1.1 Approach

Using surveys and a workshop held in Ankara, Turkey in July 2011, the INDICATE team explored the existing policy landscape and current best practice for digital preservation and examined how e-infrastructures can be used to address problems in the area of long-term digital preservation. The survey was designed to gather information about the preservation process for digitized records in a number of different countries and institutions. The topics covered by the survey included:

- Policies and strategies for global preservation, legal deposit service, copyright and access;
- What, where and how items are preserved digitally;
- Standards, administrative requirements and technical aspects;
- Back-up strategies and file integrity;
- Future planning.

Results of the survey were then analysed by the INDICATE consortium at a workshop where it was decided that the initial survey was too complex, and a simpler, more focussed survey is planned for subsequent analysis.

### 5.1.2 Examples of long-term digital preservation best practice

Existing long term preservation best practices identified

Among the institutions that completed the survey, three were identified as having effective digital preservation initiatives that have implemented best practice: the Italian National Central Libraries in Florence and Rome , the Catalunya National Library and the French National Library and Ministry. All the institutions have implemented a **detailed strategy for preservation** (e.g. specification of preferred file formats, conversion options, best storage options and system security). They have also handled most of the issues about copyrights, legal deposit services and financial sustainability. Staff members at these institutions are well trained and professional. Most of these institutions are currently **exploiting their national NREN and Grid** capabilities. **Highly available storage** has been ensured using different servers located in different places/cities and dark archives with restricted access to ensure adequate backup and recovery capabilities. **Interoperability**, at both national and international levels, has been addressed.

### 5.1.3 Best Practices

The use case study based on the initial survey of digital preservation identified a number of best practice actions, as outlined in the following sections. These best practices may be amended following analysis of the results from the simplified survey (see 5.1.1) in an update of the present Best Practices Handbook.

#### 5.1.3.1 Establish global long-term preservation strategies and policies

Currently, there is a general lack of global policies and strategies for long-term digital preservation. In most countries, long-term digital preservation is carried out by individual institutions, which have undertaken digitization and preservation using their own, individual strategy. The lack of standardized global strategies and policies makes it difficult to train staff, efficiently use digital resources and ensure financial sustainability of digital preservation initiatives. For these reasons, it is **essential to establish global/national standardization of policies and strategies for common issues related to long-term preservation**, such as training, copyright, sustainability, storage, back-up and data integrity, descriptive standards, accession, etc. The Digital Cultural Heritage Roadmap for Preservation (DCH-RP) FP7 coordinating action, expected to start in October 2012, is an example of such a strategy.

#### 5.1.3.2 Ensure long-term access to digital data

Another challenge for long-term digital preservation is long-term access to digital data. Digital technology is developing quickly and retrieval and playback technologies can quickly become obsolete. When faster, more capable and less expensive storage and processing devices are developed, older versions may be quickly replaced. When a software or decoding technology is abandoned, or a hardware device is no longer in production, records created with such technologies are at great risk of loss, simply because they are no longer accessible, resulting in **digital obsolescence**.

The Reference Model for an Open Archival Information System (OAIS) was developed in order to standardise digital preservation practice and provide a set of recommendations for preservation program implementation. The OAIS Reference Model is an ISO standard (ISO:14721:2003) and defines a common framework in order to analyse and describe concepts and terminology for Digital Archives. The major aim of the OAIS Reference Model is to facilitate a much wider understanding of what is required to preserve information in the long term, i.e. long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community.

In order to benefit from the continuing advances in software and technology, a **flexible approach to emerging data technologies and formats should be adopted** rather than mandating specific approaches. Key elements for ensuring a flexible *and* effective approach to ensure long-term access to digital data should include:

- **Virtualisation** to isolate dependencies on hardware, software and specific environments; abstract, important, interfaces/processes are identified that can be implemented on top of concrete implementations;
- **Identification of understandability gaps** to cope with changes in Knowledge Base of a community; once identified, gaps can be bridged with community knowledge before the knowledge is lost;
- **Reduction of security breaches** through systems security and data integrity to allow determination of provenance and authenticity; provenance and authenticity also depend, in part, on non-technical solutions, such as social policy, information policy and process documentation;

- **Recognition of the temporary nature of custodianship** so that techniques are implemented to allow a smooth transfer of holdings from one link in the chain of preservation to the next;
- **Provision of a definitive system of persistent actionable identifiers** which reduces the risk of deterioration of identifier systems.

#### 5.1.3.3 Ensure certified and stable digital resource reference mechanism (persistent identifiers)

The association of a Persistent Identifier (PI) to a digital resource can be used to certify its content authenticity, provenance, managing rights, and to provide an actual locator. The only guarantee of the actual persistence of identifier systems is the commitment shown by the organizations that assign, manage, and resolve the identifiers. There are a number of technological platforms in use, including:

- Document Object Identifier system (DOI [6]), widely adopted by the publishing industry;
- Archival Resource Key (ARK [7]), a URL-based persistent identification standard;
- Handle System[8], for persistent identifiers for digital objects and other resources on the Internet;
- Persistent URL (PURL[9]), a redirect-table of URLs;
- National Bibliography Number (NBN[10]), a standard identifier format currently promoted by the CENL.[11]

For DCH applications, one PI system that is demonstrating to be considered is NBN. NBN grants an exclusive namespace to national libraries so that every country has a registered sub-namespace at the Library of Congress, fulfilling the requirements of stability and permanence necessary for an institution that intends to manage a PI service. In Europe, the German and Swedish national libraries have implemented NBN registers. In Italy, the National Central Library of Florence, supported by the *Fondazione Rinascimento Digitale,* and in cooperation with the National Central Library of Rome and the Marciana National Library (Venice) is now actively working on a new NBN architecture. The NBN Persistent Identifiers system will also be implemented by the *Magazzini Digitali* project.[12]

> For DCH applications, the most appropriate PI system is NBN

The Nordic Metadata Projects[13] also uses NBN. The project aims to set up a national register of persistent identifiers for digital cultural objects on the Internet, and experiment with resolution services and access to these resources. The technology is based on the Uniform Resource Name (URN[14]), utilising the National Bibliography Number (NBN) as the reference namespace, with the aim of **facilitating citizens' access to resources, arousing sensitivity and encouraging cultural institutions to implement digital preservation programmes** for their own digital resources.

The hierarchical, multi-level, distributed approach of NBN implies that the responsibility of PI generation and resolution can be recursively delegated to lower level sub-naming authorities, each managing a portion

[6] http://www.doi.org
[7] http://tools.ietf.org/html/draft-kunze-ark-14
[8] http://www.handle.net/
[9] http://purl.oclc.org
[10] IETF RFC 3188 Using National Bibliography Numbers as Uniform Resource Names http://tools.ietf.org/html/rfc3188
[11] http://web3.nlib.ee/cenl/
[12] http://www.bncf.firenze.sbn.it/pagina.php?id=212&rigamenu=Magazzini Digitali
[13] http://www.kansalliskirjasto.fi/extra/muut/meta/
[14] IETF RFC 2141 URN Syntax http://tools.ietf.org/html/rfc2141

of the domain name space. There is a redundancy of service access points and information storage locations, **which increases the reliability of the whole infrastructure by eliminating single points of failure**. Furthermore, reliability increases as the number of joining institutions grows. A **distributed architecture also increases scalability and performance**, without altering the publishing workflows defined for the different repositories.

#### 5.1.3.4 Provide adequate training to staff involved in long-term digital preservation

The survey results indicated that there is a general lack of expertise in the area of preserving cultural heritage. Comprehensive **staff training is essential for successful digital preservation** of cultural heritage. In order to establish an effective long-term digital preservation system, training about digital preservation must include common aspects of the process, such as:

> Staff training is essential for successful digital preservation of cultural heritage.

- **Metadata:** staff should be aware of the metadata required for maintenance of a digital repository and how it is used in the context of long-term storage (metadata standards, mapping, relationships between metadata and data, metadata quality, role of metadata in internet searching, etc.);
- **Interoperability:** training should illustrate the need for interoperability and the impact it has on the use of digital archives by different users;
- **Standardization of digital archives:** staff should be aware that the digitization and conservation process in their own institution may be non-standard and the benefits of standardization should be highlighted;
- **Storage of data in multiple locations:** the importance of storing replicate data in a number of different locations (on-line backup, near-online backup, off-line backup, off-site backup) should be communicated to staff, as well as the integration of different storage media (magnetic tape, optical storage, remote backup (internet), etc.).

#### 5.1.3.5 Exploit the potential of e-infrastructures for long-term digital preservation

Long-term digital preservation faces a number of challenges, such as storage capacity, retrieval and long-term access to data, which e-infrastructures can effectively address. e-Infrastructures offer the DCH community access to:

- **High performance/throughput computing** systems to tackle large scale, computationally intensive, problems, such as document/image digital restoration;
- **High performance storage** systems to store geographically distributed replicas of file and provide fault-tolerant digital preservation;
- **Simplified authentication** systems through Single Sign-On (SSO) functionality and the support of Federations of Identity Providers;
- **Fine-grained authorization** systems to define “who” can access digital data and “what” they can do with the data (read/edit/delete data/metadata, search, browse, etc.).

While use of e-infrastructures has been widely adopted by the physical sciences, there is an under-utilisation of e-infrastructures by DCH initiatives. There are a number of challenges that will be faced as e-infrastructures are adopted by the DCH community, including unique DCH user needs, compatibility of data formats, need for interconnection of content distributed across the Grid, training, economic sustainability and a lack of standardisation in policies. However, the potential benefits of e-infrastructures to the DCH

community outlined above outweigh these challenges. Best practice will need to evolve as the e-infrastructures and DCH community address these challenges together.

## 5.2 Virtual Exhibitions

Over recent years, museums and other cultural institutions have been increasingly using the world wide web to develop virtual exhibitions and to disseminate their collections. Virtual exhibitions remove geographical barriers and allow curators to present cultural content in new ways with significant potential to engage new audiences and enhance user interactions with cultural collections. Furthermore, virtual exhibitions are not only relevant in the context of remote users, as virtual and physical exhibitions can be combined to better reach the main goals of cultural institutions (research, dissemination, communication, learning and enhanced experiences for cultural users).

### 5.2.1 Approach

The Virtual Exhibition Case Study (D5.2) examined the current landscape of virtual exhibitions and how e-Infrastructures can be used to create new projects, using virtual performances (performing arts actions and experiences that use the possibilities of Grid Computing, multimedia and interactive technologies and virtual spaces) as an example. A survey on virtual exhibitions in the INDICATE partner countries was carried out and the results were analyzed at a workshop in Jordan in December 2011. Additionally, papers, reports, articles, wiki platforms and other resources on virtual exhibitions were analyzed to provide a global overview on virtual exhibitions technologies and their use by the cultural institutions. An important contribution to the case study was the guidelines for online virtual exhibitions developed by an Italian national working group in 2010.[15]

E-Infrastructures offer virtual exhibitions new formats, more storage and greater scalability

### 5.2.2 Results

The results of the use case study show that museums and other cultural institutions are using virtual exhibitions and virtual performances to achieve their institutional aims (research, education, dissemination, etc.) and reach new audiences. The use case study presented a range of example projects, as well as background information about the common technological aspects of virtual exhibitions and performances.

One of the key findings was that the capabilities of e-Infrastructures for virtual exhibitions and performances are not widely exploited. However, with their superior computing capacity**, e-infrastructures can be used to generate new virtual exhibition formats** that would not be possible with traditional computing approaches. For example, the ASTRA project is using the GEANT network and Grid infrastructure to reconstruct the sound of ancient instruments that no longer exist using archaeological data. The physical modelling required is computationally intensive and so benefits from the computational power of Grid infrastructure. E-infrastructures also offer greater **scalability and storage capacity** for virtual exhibitions than traditional approaches.

[15] Mostre virtuali online: linee guida per la realizzazione, Roma: Ministero per i beni e le attività culturali, http://www.otebac.it/index.php?it/320/mostre-virtuali-online-linee-guida-per-la-realizzazione

### 5.2.3 Best Practices

#### 5.2.3.1 Use virtual exhibitions and performances to eliminate barriers and expand audiences

By nature, virtual exhibitions and performances can be used to eliminate barriers between the DCH content of museums and cultural institutions and their intended audiences. Virtual exhibitions and performances can be designed to **allow access regardless of the physical location** of the viewer. In certain cases, geographical barriers due to the different physical locations of different cultural items related to a common theme can be overcome by virtual exhibitions. Also, **virtual exhibitions offer benefits to disadvantaged groups**, such as people with mobility or visual/auditory impairments, by allowing them the opportunity to enjoy cultural content that they may be unable to access otherwise. Wherever possible, **multilingual virtual exhibitions** should be used to reach as wide a user base as possible.

#### 5.2.3.2 Stimulate user involvement/interaction with virtual exhibitions

One of the strengths of virtual exhibitions is that they allow greater **interaction** between the user and the cultural material being presented. This interaction should be facilitated and enhanced wherever possible **by using techniques to engage and empower the exhibition users**, such as strategic use of hypertext and multimedia. Exploiting hypertext to provide numerous links between information units **allows the user to direct their own visit**. In some cases, involvement of the user as a supplier of content can enhance interactivity and enrich the exhibition.

#### 5.2.3.3 Consider technological aspects of virtual exhibitions and performances

There are a number of technological aspects that should be considered during the design of virtual exhibitions and performances to ensure that they effectively deliver their intended cultural message and persist in time after their initial launch. Exhibitions and performances should be served by the technology used to deliver them and not the reverse. **Scalable system architecture** should be used in the design of exhibitions to enable enrichment of online content and services over time. Indeed, **re-usable technology** should be used wherever possible and **inter-operability** should be ensured so that exhibitions do not become obsolete or without meaning over time.

#### 5.2.3.4 Respect copyright rules when preparing virtual exhibitions

When preparing virtual exhibitions, **copyright rules must be respected**. The copyright status of each digital item that will be used in the exhibition must be determined. Items that are under copyright may be used under waiver from the copyright owner or for a fee. The **model for re-use of content** from the exhibition, whether open (such as some Creative Commons[16] licenses) or closed (copyright), must be decided for both the individual digital items used and the overall exhibition.

#### 5.2.3.5 Use virtual exhibitions to drive commercial applications of DCH

Virtual exhibitions and virtual performances attract and interact with new audiences, and can have a significant impact on the commercial applications of DCH, such as **cultural tourism, creative industries and merchandising**. For example, in Genoa, QR codes (two-dimensional bar codes that can contain data or links) have been used on monuments around the city for tourism in support of the Festival of Science. [17]

[16] http://creativecommons.org/

[17] http://www.viveregenova.comune.genova.it/content/festival-della-scienza-qr-coded-citta-digitale-verso-la-realta-aumentata

#### 5.2.3.6 Exploit e-infrastructure capabilities for virtual exhibitions and virtual performances

Currently, virtual exhibitions and performances by the DCH community are not fully exploiting the capabilities of e-infrastructures. However, e-infrastructures provide high capacity services that cultural institutions cannot afford in-house. The cultural virtual exhibitions and performances sectors should engage with the e-infrastructures sector to **develop more innovative and complex technologies beyond the traditional computation, storage and connectivity services** currently offered by e-infrastructure providers. Innovation in this domain could include on-demand 3D-rendering on demand, video trans-codification and 3D image reconstruction, and other software modules generally installed in-house will extend the boundaries of what is possible.

## 5.3 Geo-coded Digital Cultural Content (GCC)

Geocoding digital cultural content adds geographic location information (typically latitude and longitude) to the content. Geographic location information can provide a range of information, such as a description of the provenance of an item, the current location of a cultural object or the location of an event related to a cultural object. **Geocoding adds value to digital cultural content** in a number of ways, including:

- Allowing users to efficiently browse through the content of cultural portals using space and time;
- Supporting content mapping;
- Allowing GIS calculations and simulations to be performed;
- Overlapping architectural/archeological heritage with museum objects and intangible heritage;
- Defining the protected areas of monuments;
- Allowing geo-visualizations and historical simulations;
- Offering new functionality for cultural tourism, education, etc.

GCC provides information that is useful in DCH research, restoration, conservation and preservation, and can also be used for the promotion of cultural heritage through digital libraries and cultural institution portals. There are also applications for GCC in the context of cultural tourism and re-use of cultural content by the creative industries. The processing power and storage capabilities of e-infrastructures can be efficiently and cost-effectively exploited by all these applications.

### 5.3.1 Approach

A workshop was held in Ljubljana, Slovenia in Feb 2012 to carry out the GCC use case study. The workshop included the presentation of best practice examples of the use of GIS technology for access to digital cultural content, discussions of how e-infrastructures can be used to enhance GIS cultural applications and an analysis of the resources offered by e-infrastructures offer and how these resources can be deployed for GIS cultural implementations. Case studies related to the use of geo-coded digital cultural content from Italy, Ireland, Egypt, France, Slovenia and Jordan were presented. A world café group activity was also used to facilitate discussion of archaeological and architectural heritage and GIS, Libraries and GIS, and Museums and Cloud Computing by the INDICATE partners. The findings of this workshop formed the basis of the GCC use case study report, detailing the current approaches and emerging research

The cultural virtual exhibitions and performances sectors should engage with the e-infrastructures sector .

and development in the area of geocoding of cultural content in digital libraries, cultural tourism, heritage, e-learning, and other cultural areas, focusing in particular on the possibilities and benefits of using e-infrastructures (cloud and Grid computing and data infrastructures). Also included in the use case study report were the results of testing of geo-parsing and geo-tagging e-services for digital culture.

### 5.3.2 Results

Due to the computational complexity of the tasks, e-Infrastructures are particularly useful when geo-referencing and geocoding large historical maps, when geocoding precise locations in wider geographical areas and when geo-parsing and assigning large numbers of textual digital objects. Furthermore, linking open geographical data requires a stable environment that is always accessible, which e-Infrastructures can provide at relatively low cost.

Testing of the Europeana Geo-parsing service by INDICATE showed that geo-parsing can be effectively used to automatically assign geographic coordinates and that the current service is useful for geographic details from the level of large towns, regions and upwards. Greater spatial accuracy, provided from sources such as databases of local geographic names, will improve future geo-parsing outputs.

### 5.3.3 Best Practices

#### 5.3.3.1 Adopt standards from the digital geographic information field

It is essential that the DCH community adopts standards from the digital geographic information field to ensure the assignment of accurate, consistent, usable geographic information for digital cultural content. Best practice standards should be derived from the following four sources:

- The **Technical Committee 211 of the International Organization for Standardization** (ISO/TC 211 Geographic information/Geomatics) is responsible for standardization in the field of digital geographic information. In February 2011, the ISO published 52 standards under the direct responsibility of the Technical Committee 211. The standards address topics such as infrastructure for geospatial standardization, data models for geographic information, and encoding of geographic information and specific thematic areas. The ISO/TC 211 group of standards provides for a **fundamental structure for geographic information that enables its use in computational processing**;
- The **Open Geospatial Consortium** (OGC) is a voluntary consensus organization that is closely linked with other international standardization bodies, such as the ISO/TC 211. In fact, the OGC standards Web Map Service, Geography Markup Language (GML), and Simple feature access have become ISO standards. The OGC is leading the development of standards for geospatial content and location-based services and also for GIS data processing and sharing. The organization encourages the **development and implementation of open standards, free and openly available to the market**. Most importantly, the OGC aims to **ensure interoperability in the processing of geographical information**, which is key for designing information systems in the cultural heritage field;
- The **Infrastructure for Spatial Information in the European Community (INSPIRE) Directive** aims to ensure compatibility and usability of spatial data infrastructures of Member States in the European Union. The Directive came into force on 15 May 2007 and is being implemented in stages until completion in 2019. The Directive requires that common Implementing Rules be adopted in a

number of areas, such as metadata and data specifications. These Implementing Rules are adopted as Commission Decisions or Regulations, and are binding in their entirety. The **INSPIRE spatial infrastructure will be useful in the digital cultural heritage field** as the Implementing Rules governing coordinate reference systems, geographical names and administrative units can be used as both a methodological background and technical standards. Also, INSPIRE spatial data (ortho-imagery and geographical names) could be used for representing digital cultural content on web maps;

- The **European Terrestrial Reference System 1989** (ETRS89) should be used as the coordinate system for geocoding cultural content, especially protected architectural and archeological sites. While the World Geodetic System 1984 (WGS84) geographic coordinate system is the most widely used, it is not spatially accurate enough. **ETRS89 is used as the standard precise GPS coordinate system throughout Europe** and is linked to the European continent.

#### 5.3.3.2 Use e-infrastructures for GCC of architectural and archaeological digital heritage

Geographic information systems have long been used in the fields of architectural and archaeological heritage. e-Infrastructures can **cost-effectively provide the computational capacity for innovative applications,** such as 3D modeling for research, restoration of monuments and 3D (or 4D) presentation of cultural heritage to the general public. Grid computing is also well suited to risk scenario simulations and risk management for cultural heritage, spatial analysis and geoprocessing services.

#### 5.3.3.3 Use e-infrastructures to exploit GCC for cultural tourism

Cultural tourists are interested in the culture of the geographical areas they visit, including aspects such as art, history and architecture. Furthermore, a recent OECD report has indicated that cultural tourism has the potential to influence regional development.[18] At present, tourism portals, such as www.booking.com, www.hotels.com and www.lonelyplanet.com, have interactive geographic maps with information about accommodation, transport and other places of interest. However, there is minimal information about cultural heritage available through these portals. To more fully develop the potential of cultural tourism, **cultural information from cultural portals should be integrated with tourism portals**. To achieve this, a stable and reliable interoperability framework is needed, which could be effectively provided using cloud based systems. Grid computing should also be considered for more computationally intense applications, such as calculating the shortest path between cultural locations of interest.

The **significant processing power of Grid computing is well suited for geo-parsing**

#### 5.3.3.4 Use Grid computing for geo-parsing

**Geo-parsing can be used to enrich digital cultural items** by automatically assigning geographic coordinates to textual words and phrases so that geographic coordinates can be assigned to cultural objects where there are no available coordinates or additional coordinates are required. Geo-parsed features can then be mapped and entered into a geographic information system. The **significant processing power of Grid computing is well suited for geo-parsing**, which relies on computationally complex processes, such as natural language processing, pattern recognition and web semantics.

[18] OECD (2009) The Impact of Culture on Tourism. OECD, Paris

# 6 Research Pilot and Use Case Study Best Practices

As demonstrated by the results of the preceding sections, research pilots and use case studies are effective research methods for validating theoretical approaches and gaining new knowledge, and can be applied to research in the area of DCH.

## 6.1 Use research pilots to test new approaches in DCH

Research pilots are an effective way to test the applicability and usefulness of innovative applications and technologies for use by the DCH community on a small scale before large-scale deployment. The experiences obtained from the research pilots can be used to tailor applications for a better fit with the needs of the DCH community and even identify technologies and applications that are not suitable for DCH. The costs and risks associated with research pilots are low, but the potential benefits, such as identifying potential shortcomings of a new technology before full-scale implementation, are significant.

## 6.2 Use focused surveys and workshops for use case studies in DCH

The effectiveness of use case studies is dependent on the quality and completeness of information gathered about the cases under study. An effective way to gather case information is to prepare focused surveys that are distributed to a network of experts from a range of expertise and locations. Questions posed in the surveys should be specific and structured enough to ensure that common information about all cases is obtained so as to allow useful comparisons, but also allow some scope for respondents to contribute additional, perhaps unexpected, information that can reveal unique information previously unknown to the surveyors. Analysis of the survey responses from experts also requires the input of experts. For this reason, workshops where the survey results are discussed and analysed by a network of experts are especially effective.

# 7 Conclusions

In this handbook, best practices for the use of e-infrastructures by the DCH community have been described based on the results of the INDICATE research pilots and use case studies. In line with the aims of the INDICATE project, these pilots and use case studies were designed to practically test and validate the policies and best practice identified by the DC-NET sister project. While the use of e-infrastructures by the DCH community is increasing, the capabilities of e-Infrastructures have yet to be fully exploited. Adoption of the best practices outlined in this handbook will ensure that:

- Future DCH activities, such as long-term preservation, GCC and virtual exhibitions, will **fully exploit the capabilities of e-infrastructures** (scalability, processing power, storage capacity) for computationally expensive processes;
- The DCH and e-infrastructures communities will work together to **create innovative applications of e-infrastructures** that will be applicable both for DCH and other areas;
- Barriers to engagement of users with DCH, such as geographical location, language and the sheer volume of digital content, will be minimized or even eliminated;
- Access to and use of e-infrastructures by non-IT experts will be facilitated to **broaden and deepen engagement with DCH collections** by a wider audience.

Furthermore, the results shown here support the conclusion that research pilot studies and research use case studies are effective methods for identifying best practice in the area of DCH.

The Handbook has been presented at the Final Conference in Ankara, printed and made available on the website and it is one of the instruments created in the INDICATE project that will last beyond the EC funding period. All the partners are in fact committed to continue the distribution of the Handbook and its promotion.

# 8 Glossary

**Application Programming Interface (API)** - technical interface used for searching and retrieving metadata.

**ASTRA -** Ancient instruments Sound/Timbre Reconstruction Application project, http://www.astraproject.org/

**e-Infrastructure -** an environment where research resources (hardware, software and content) can be readily shared and accessed where necessary to promote better and more effective research.[19]

**ETRS89** - European Terrestrial Reference System 1989

**GCC -** geo-coded digital cultural content

**gLibrary** - a digital management asset system for the Grid.

**Identity Federations** - made of *"[…] the agreements, standards, and technologies that make identity and entitlements portable across autonomous domains (Burton Group)*". Identity Federations have the aim of setting up and supporting a common framework for different organisations to manage accesses to on-line resources.

**INSPIRE -** Infrastructure for Spatial Information in the European Community

**ISO/TC211 -** Geographic Information/Geomatics: Technical Committee 211 of International Organization for Standardization standards

**MICHAEL -** Multilingual Inventory of Cultural Heritage in Europe, project funded by the European Commission's eTen programme

**RDF -** Resource Description Framework, a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model; used as a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats.[20]

**Science Gateway** - "*community-developed set of tools, applications, and data that is integrated via a portal or a suite of applications, usually in a graphical user interface, that is further customized to meet the needs of a specific community (US Teragrid project).*"

**Virtual exhibitions** - *is a hypermedia collection made up of digital items which are:*

- Linked together by a common threat, an inter-disciplinary topic, a concept, an idea, an
- anniversary, a special event, or a physical person;
- Displayed in 2D and/or 3D;
- Occasionally stored in distributed networks;
- Made accessible through the potential provided by modern technologies, thanks to a system architecture designed to provide user-centred, absorbing experiences;
- Dynamic products that can offer services and be updated periodically.

*A collection of digital items, in and of itself, does not constitute a virtual exhibition. It is only when the items are carefully selected to illustrate a topic, and are tied together forming a narrative or a logical itinerary, that they constitute an exhibition.* [21]

[19] http://cordis.europa.eu/ictresults/index.cfm?ID=90825&section=news&tpl=article
[20] http://en.wikipedia.org/wiki/Resource_Description_Framework

**Virtual performances -** *Performing arts actions and experiences that use the possibilities of Grid Computing, multimedia and interactive technologies and virtual spaces over the network.*[21]

**WGS84** - World geodetic system dating from 1984

**XML -** Extensible Markup Language, a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.[22]

[21] INDICATE D5.2
[22] http://en.wikipedia.org/wiki/Xml