**Project Acronym:**  DCH-RP

**Project Title:** Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for DCH in 2020

**Contract Number:** 312274

**Starting date:** 1st October 2012 **Ending date:** 30th September2014

**Deliverable Number:** D 5.1

**Title of the Deliverable:** Technical Plan for DCH-RP proofs of concept

**Task/WP related to the Deliverable:** WP 5, Task 5.1

**Dissemination Level:** Public

**Author(s):**  Michel Drescher, EGI.eu

**Partner(s) Contributing:** All WP5 partners

**Contractual Date of Delivery to EC:** Month 3

**Actual Date of Delivery to EC:** Month YYYY

**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. 06 4959302

*E-mail :* rcaffo@beniculturali.it

*Project WEB site address :* http://www.dch-rp.eu

**Context**

|  |  |
| --- | --- |
| WP 5 | Proofs of concept |
| WP Leader | EGI.eu |
| Task 5.1 | Technical planning of the first run of proofs of concept |
| Task Leader | EGI.eu |
| Dependencies | WP3 |
| Starting date | 1st October 2012 |
| Release date | dd mmmm yyyy |

|  |  |
| --- | --- |
| Author(s) | Michel Drescher (EGI.eu) |
| Contributor(s) | Rosette Vandenbrouke (BELSPO), Sanja Halling (RA), Maciej Brzezniak (PSNC), Roberto Barbera (INFN) |
| Reviewers |  |
| Approved by: |  |

**History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Author** | **Comments** |
| 0.1 | 23 Oct ‘12 | Steve Brewer (EGI.eu) | Initial version |
| 0.2 |  | Steve Brewer (EGI.eu) | Initial structure of document |
| 0.3 | 15 Jan ‘13 | Michel Drescher (EGI.eu) | Document template, updated structure, section reviews, Introduction section |
| 0.4 | 17 Jan ‘13 | Michel Drescher (EGI.eu) | Expanded on section 3.1, 3.2, added Swedish contribution (section 4.2) |
| 0.5 | 18 Jan ‘13 | Michel Drescher (EGI.eu) | Integrated BELSPO contribution, integrated RA contribution, further process in section 3 |

Table of Contents

1 EXECUTIVE SUMMARY 4

2 Introduction 5

2.1 Objectives of the deliverable 5

2.2 Structure of the document 6

3 Methodology, Tools and Processes 7

3.1 Audience and stakeholders 7

3.2 Agile Project Management vs Waterfall model 8

3.3 Timetable for activities 9

3.3.1 Contractual work packages deliverables and milestones 9

3.4 Toolchain and documentation 10

3.5 Iterations and Continuous testing 11

4 Proofs of Concept (PoC) 12

4.1 Italy (INFN, ICCU) 12

4.2 Sweden (RA) 12

4.2.1 Pilot lead partner 12

4.2.2 Environment and constraints of the pilot 12

4.2.3 Objectives 13

4.3 Belgium (BELSPO) 14

4.3.1 Pilot lead partner 14

4.3.1 Environment and constraints of the pilot 14

4.3.2 Objectives 14

4.4 Estonia (EVKM) 14

4.5 Poland (PSNC) 14

4.5.1 Pilot lead partner 14

4.5.2 Environment and constraints of the pilot 14

4.5.3 Objectives 15

4.6 Hungary (NIIFI) 15

5 Conclusion 16

6 Annex A – SCRUM: Agile Project Management 17

# EXECUTIVE SUMMARY

# Introduction

Work Package 5 has been designed to coordinate and carry out a number of proofs of concept (PoCs) using a number of pilot deployments allowing end users in the Digital Cultural Heritage community to store, access and manage digitalized facsimiles of artifacts of human culture using contemporary distributed IT infrastructures. While the coordination of the piloting work is a necessary part of this work package, the true core of this Work Package is to conduct the proofs of concept, and its main outcome are the results and lessons learned that are produced by these proofs of concept.

These PoCs will be influenced by two major sources of input: Work Package 3 (Preservation Roadmap) formulates milestones that need to be achieved and technical requirements that must be satisfied to progress along the preservation roadmap over the next couple of years. The areas of interest of the Preservation Roadmap will be, among others:

1. What are the short, medium and long-term milestones for preservation infrastructure through 2014, 2016, 2018 for the following?
   1. Harmonisation of data storage and preservation?
   2. Progress for inter-organisational communication?
   3. Establishment of conditions for cross-sector integration?
   4. Governance models for infrastructure integration?
2. Standards – what impact do emerging and established standards have on the intended solutions
3. Registry of tools and services – how useful is the pilot registry
4. Trust building activities – to what degree do the pilot systems contribute to trust building?
5. Best practices, per country, per domain, per technology
6. Interoperability – see MINERVA project for examples. The proof of concept trials should address the following issues at least to some degree:
   1. Technical
   2. Semantic
   3. Political/human
   4. Inter-community
   5. Legal
   6. Internationality

While WP3 defines the common technical parameters and, to a certain degree, the goals of the PoCs (“*How* is it contacted?”) WP4 (Case Studies and Best Practice) complements this by providing the context, material, and workflows for the various domains within the DCH community (“*What* needs to be processed and how do users do that?”).

For a large part, these Proof of Concepts share many principles and best practices with software quality assurance and software engineering activities that are regularly carried out before any piece of software is released for general availability. The coordination aspect of this work package introduces suitable techniques and tools, processes and guidelines in order to ensure that the work carried out as part of the PoCs are repeatable, auditable, and suitable as input into the preservation roadmap developed in Work Package 3.

## Objectives of the deliverable

This document provides the main output of Task 5.1 and captures the planning work for the PoCs to take place over the next phases of the project. The scope and actual content of this document are limited by design; rather, it is designed as a “book of guidelines”, processes and tools that the actual proofs of concept activities should follow. This is because the overall planning work covers both shared and disparate aspects of technical activities: The partners in Work Package 5 conducting the pilot work will do so using domain specific data and workflows (e.g. musea with digitalized paintings, digitalized 3D objects, contemporary digitalized art, films, literature archivists, music archivists, anthropologists, etc.), and thus are likely to use domain specific data formats, metadata descriptions, and many more different aspects. The planning pertaining to domain specific activities within the Proof of Concepts are left with the respective partners and will be informed mainly by the input coming from Work Package 4. This domain specific planning will be captured elsewhere as it is much more dynamic in nature and requires a different approach than a fixed document against which the activities are conducted.

Instead, this document will capture the *common structure and processes* with which the PoCs will be planned, conducted and documented. Mainly informed by the input from Work Package 3, Deliverable 5.1 establishes the processes by which the experiments and tests ensure that the technical requirements of data preservation will be covered and results documented.

## Structure of the document

The remainder of this document is divided into two parts and an annex as follows.

Section 3 describes in sufficient detail the common processes, tools and methodology of how the Proofs of Concept will conduct their work. Based on established agile project management methodology, a highly iterative approach has been chosen to establish a continuous evolution and improvement of the pilot activities. The idea behind this approach is that while the end-goal is clear (a complete roadmap for DCH preservation in the coming decades) a highly iterative approach allows gathering and documenting first results early on for public dissemination and information, and thus allowing for intervention and changing direction very quickly if necessary without significant loss of effort.

Section 4 describes the individual Proofs of Concepts on a high level, which specific domain is targeted, which tools are tested and for what reason, and the overarching goals that these specific pilot activities wish to reach. This section does not constitute a complete planning document for the Proofs of Concept; it is included in this document to ensure that there is sufficient spread of activities across DCH domains, and necessary but not excessive overlap in effort to be able to identify tools and services that may be re-used across domains.

This document concludes with a summary of the identified plans, and gaps, and the next immediate actions to progress towards practical experiments with the envisioned Proofs of concept.

Attached to this document, Annex A provides additional background information on the Agile Project Management methodology that forms the basis of the processes established in this Work Package.

# Methodology, Tools and Processes

## Audience and stakeholders

The DCH-RP project aims to play an important role in taking forwards the adoption of e-Infrastructure as a computing platform by the DCH research community. The following table describes the stakeholders for technical plan and its outputs. The term stakeholder in this context refers to the audience and participants of the proofs of concept experiments.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Stakeholder | Participant/ Observer | Role | Area of interest | Importance of information |
| EGI.eu | P | Task leader | All | High |
| INFN eCSG | P | Science Gateway provider | Interfaces, usage, requirements | High |
| WP4 | P | Provider of Case Studies and Best Practice | Functionality of experiments, usability results | Medium |
| WP3 | P | Developer of Roadmap | All aspects of results as defined in scope | High |
| WP2 | O | Dissemination and sustainability | Success stories and lessons learned from tests, validity of results | High |
| WP1 | O | Project Management | Effectiveness and success of the tests | High |
| Cultural Heritage institutions | O | Prospective partners in the services and end users of the services | Usefulness and usability of the new services, security and reliability of new services, interoperability | Medium |
| Cultural Heritage services end-user community | O | Prospective users of the new services | Usefulness and usability of the new services, range of services | Medium |
| e-Infrastructure providers | O | Potential hosts for these new services | Applicability of new services, usability, ease of configuration, scalability | Medium |
| General public | O | Potential users, general interest in topic | Awareness that the domain of digital cultural heritage is evolving swiftly | Low |

Table 1: Stakeholder of the WP4 activities

This diverse set of stakeholders calls for a very open and flexible way of conducting the activities in Work Package 5.

The diversity in interest in the work package activities requires different approaches in communicating the results. Stakeholders with relatively low interest in the proceedings of Work Package 5 are expected to be interested in comparatively infrequent but complete reports provided in referencable material, such as papers, or documents, project deliverables and milestones. On the other end of the spectrum, stakeholders with high interest are likely to appreciate frequent updates of anything that happens within this work package, even if it is incomplete or incremental information.

The sheer number of stakeholders indicates that we can expect a good deal of communication and dissemination. Although Work Package 2 is dedicated to project wide activities in this direction, the effort within WP5 dedicated to communication and dissemination should be minimized as much as possible. Choosing the right set of tools that are able to support and automate interactions between work packages, between activities is essential. Essentially, the success of this work package is strongly influenced by the *communication and collaboration between people*; collaboration tools and methodology should satisfy this prime requirement.

## Agile Project Management vs Waterfall model

Classic project management organizes work in well-defined sequential phases. Commonly known as the “waterfall model”[[1]](#footnote-1) this methodology assumes a well-known and complete understanding of the problem space before any work is undertaken. Derived mostly from manufacturing processes, the waterfall model assumes that work linearly progresses until the planned outcome is achieved. This requires meticulous preparation, documentation and collection of requirements and specifications against which the process will produce its result.

In the real world however, particularly in science projects of exploratory nature such as the DCH-RP project, not all preconditions and requirements of the overarching objectives are known in the beginning. Such environments call for a much more resilient and flexible way of project management. In general, all agile project management methodologies share the fundamental concept of feedback loops in iterative cycles of activities. The idea behind this model is that high-frequent iterations ending in feedback activities allow for quick interventions and corrective measures where required, and to adjust the direction of the project or some of its activities.

This paying respect to and embracing the unknown at the beginning that is common among agile methodologies is recognizable on all ends. Frequent iterations are only one aspect; concepts such as retrospective assessment, “planning poker”[[2]](#footnote-2), regular “stand up meetings”, “user stories”[[3]](#footnote-3) and product backlogs all pay attention to:

* Knowledge must be shared among all members of the agile managed activity
* Regular supervision of the past iteration and actions for process improvement
* Stakeholder satisfaction
* Use of domain language as much as possible (contributes also to stakeholder satisfaction)
* Diversity in participants skill set is a benefit, not a drawback

Considering the diversity of members in this project, and particularly in Work Package 5 agile activity management promises the effectiveness and result orientation that is necessary for the success of the project.

## Timetable for activities

The detailed time planning needs to keep into account the preparatory planning, necessary technology activities and dependencies. As most of the PoCs will include the e-Cultural Science Gateway (eCSG) in their activities (see section 4 for more detail) the proper upgrade and deployment of this portal is a significant dependency for the progress of this Work Package.

Detailed technical planning for the Proofs of Concepts must take this into account, as well as potential risks pertaining to availability of cultural data, as well as the e-Infrastructure to work with the cultural data.

### Contractual work packages deliverables and milestones

The following milestones and deliverables are contractually agreed and as such non-negotiable elements of the work plan:

* D5.1 – Technical Plan (M3)
* D5.2 – Upgraded eCulture Science Gateway (M6)
* D5.3 – Report on first Proof of Concept (M12)
* D5.4 – Report on the second Proof of Concept (M21)
* MS12 – Technical planning (M3)
* MS13 – eCulture Science Gateway upgraded (M6)
* MS14 – First Proof of Concept completed (M8)
* MS15 – Second Proof of Concept completed (M15)

The deliverables constitute the formal and final written records of Work Package 5. The corresponding milestones are designed as points in time by when the described activities are planned to conclude; the results of such activities then feed into the preparation and finalization of the respective deliverable. It is for this reason that most of the milestones are timed in advance of the formal project deliverables.

Therefore, the milestones can be relaxed if in turn the deliverables’ deadlines are carefully observed. In fact, when applying agile activity management techniques, the effort of running the proof of concepts, documenting tasks and results is leveled out much more evenly over the whole duration of the planned Proofs of Concept periods while increasing communication between stakeholders and delivering a constant stream of results. The following table provides an overview of the planned phases of the activities within Work Package 5.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Month** | **M1** | **M2** | **M3** | **M4** | **M5** | **M6** | **M7** | **M8** | **M9** | **M10** | **M11** | **M12** |
| Activity | Agile iteration planning and facilitation | | | | | | | | | | | |
| Technical plan |  |  | D5.1 |  |  |  |  |  |  |  |  |  |
| Activity |  |  |  | Proof of Concepts without eCSG | | | | | | | |  |
| Upgraded eCSG |  |  |  |  |  | D5.2 |  |  |  |  |  |  |
| Activity |  |  |  |  |  |  | Proof of Concepts using eCSG | | | | |  |
| Prepare first report |  |  |  |  |  |  |  |  |  |  |  |  |
| First PoC report |  |  |  |  |  |  |  |  |  |  |  | D5.3 |
| **Month** | **M13** | **M14** | **M15** | **M16** | **M17** | **M18** | **M19** | **M20** | **M21** | **M22** | **M23** | **M24** |
| Activity | Agile iteration planning and facilitation | | | | | | | | |  |  |  |
| Activity | Proof of Concepts without eCSG | | | | | | | |  |  |  |  |
| Activity | Proof of Concepts using eCSG | | | | | | | |  |  |  |  |
| Prepare second report |  |  |  |  |  |  |  |  | |  |  |  |
| Second PoC report |  |  |  |  |  |  |  |  | D5.4 |  |  |  |

Table 2: Chronological planning of WP5 activities

## Toolchain and documentation

Collaborative tools

* Wiki
* Document database
* Single sign-on
* Meeting planner
* Discussion forum
* Task manager

Agile activity management artefacts

* User stories
* Requirements
* Backlog
* Iteration & Team Velocity
* Tasks
* Stand-ups and peer-reviewng
* Reporting

## 

This provides a rough outline of the first steps common to all Proof of Concepts, applying the tools and techniques described in section 3.4

# Proofs of Concept (PoC)

*<<This is the section where we need input from the partners conducting the Proofs of Concepts. Some input was already sent, which is obviously not integrated yet.*

*The objective of the input are:*

* *Provide the specific PoC objective*
* *The scientific domain (see introduction for examples)*
* *Envisioned tools and technical architecture of the overall tool chain*
* *Outline description of the external infrastructure that is expected to be used (as opposed of providing it directly)>>*

## Italy (INFN, ICCU)

*<<Contribution was sent, but is pending integration>>*

## Sweden (RA)

### Pilot lead partner

**Digisam** is a secretariat for National coordination of digitisation, digital preservation and digital access to cultural heritage. In order to coordinate the continued development work on digitisation issues, and to coordinate the activities connected to the National Digital strategy within the timeframe of 2012-2015, the government has established a coordinating secretariat for digitisation, digital preservation and digital access to the cultural heritage – Digisam. Digisam started its work in the autumn of 2011 and is organized as a department at the National Archives of Sweden. The main task is to promote the achievement of the objectives of the national strategy for digitisation.

**Riksarkivet** is the formal the DCH-RP project partner; Digisam, established by and with a direct task from the Sewdish Government, is a first level department in Riksarkivet and as such will lead the contracted Proof of Concept activities within Work Package 5.

### Environment and constraints of the pilot

According to the Description of Work the Swedish partner in the project, the National Archives will organise concrete experiments of the concepts established by WP3, with e-infrastructure facilities provided by the NGI. After the first plenary meeting of DCH-RP where project partners discussed on which kind of DCH material would be interesting to use for the experiments, Digisam has been in contact with following cultural heritage institutions that all showed interest in participating in experiments with their data:

* The National Archives
* The Museums of World Culture
* The Swedish Museum of Natural History
* The Authority of the three museums:
  + The Royal Armoury,
  + Skokloster Castle
  + The Hallwyl Museum
* Swedish National Museum of Science and Technology
* National Maritime Museums
* The Institute for Language and Folklore
* The Swedish museum of Architecture

In the cultural heritage sector there are today different preservation requirements. Currently, there are often no specific routines and support processes for the management of cultural heritage information in digital form in order to prepare it for long-term preservation. Often there is also a lack of resources for the development and management in terms of procedures for creation of digital management, procedures for selection processes, and quality assurance. In practice, storage solutions are often only technical storage, which is short-termed, without any authenticity or preservation of context. Systems that are used for management of the information such as museum system cannot, in most cases, provide long-term preservation functionality. Joint processes through e-infrastructures could mean a lower cost and higher quality than if produced at each individual authority which also contributes to an increased availability of digital information.

Within CH institutions, databases, files are often stored in several sizes. If you take for example images - they are often stored in multiple copies in the CH institution database but also saved on a separate server in a high-resolution format. In addition, raw-data files are stored on a local server. If such as comprehensive and complete material as possible should be tested (including image files in high resolution and raw-data file format, but also texts, pre-listings and links), it probably means additional technical work for system administrators of the database.

The most important pre-condition for the tests is to clarify what the institutions benefits from this solution – what are the functions we will test that do not exist today and if those functionalities are already available/used, what can be improved or more efficient. This is also a crucial issue for selection of data for tests.

All of the interested CH institutions are connected to Internet through the Swedish e-Infrastructure provider/NREN, SUNET. However, in order to estimate how many resources are needed from the institutions to be able to deliver data in e-CSG and to test the preservation functions we need to describe more specifically the data amount and path from the institutional database(s) or other storage solutions to the proposed e-infrastructure and what it means from for example security aspect.

### Objectives

General guidelines, proposals for division of responsibility and how an integrated digital information management and a coordinated and cost-effective digital preservation should be designed and for making information accessible and usable in digital environments are key issues for Digisam to handle.

Digisam will contribute to a proposal for national guidelines for an integrated digital information management and a coordinated and cost-effective digital long-term preservation of collections and archives, including audiovisual archives, can be done at the state institutions that collect, preserve and make available cultural material and cultural information. A central issue is finding common and cost effective solutions for long-term digital preservation of common standards for metadata is a central and critical issue for achieving the overall goal.

The proposal should also include a role and responsibility for the work of aggregation, making available digital and digital preservation within the state's cultural heritage sector and highlight the needs and conditions for the use of common and cost effective solutions for long-term digital preservation of common standards for state authorities as well as the basic requirements that are necessary in a common basic infrastructure and services which can facilitate this process and be developed further.

## Belgium (BELSPO)

### Pilot lead partner

Belspo, the Belgian Science Policy, is the project partner that is responsible in WP5 for the organisation of the proof of Concepts. Belspo is not a cultural heritage itself but reaches out to federal cultural heritage institutions for obtaining data, taking up tools from WP3 to test and check out the concepts defined in the roadmap. The four cultural showed their interest in participating actively in the project: the Royal Institute for Arts (KIK), the Royal Museum for Arts and History (KMKG), the Royal Library (KB) and the State Archives (RA). The data of the KIK and KMSG are of the same type while the KB and the RA have didgital documents.

### Environment and constraints of the pilot

The above mentioned cultural institutions already have experience in archiving their digital data. They will make part of this archived data available via the e-infrastructure environment provided by the Belgian National Grid Infrastructure (BEgrid). They are already all connected to the Belgian research network.

The cultural institutes wish to use the formats and structure of the archiving method they have chosen. As they took care of following established standards in the field they hope that interoperability will not be a problem.

Problems could arise when the e-CSG would show not to be usable on the BEgrid infrastructure. Problems to be solved are the attainment of robot certificates for e-CSG, the membership of the Belgian authentication federation for those institutes or for the DCH community.

IPR problems could also arise but this is a topic that is included in the DoW.

### Objectives

Key issues for the Belgian participants is testing of real life data and situations so that on a positive result of the PoC the basis is laid for a sustainable DCH data infrastructure and corresponding services. They also expect that the cooperation with the e-infrastructures will result in novel solutions for the long-term preservation of their data. The common definition of tools and services should also accelerate a still better exploitation of the archived data.

## Estonia (EVKM)

## Poland (PSNC)

### Pilot lead partner

PSNC

### Environment and constraints of the pilot

The Polish CH community currently uses a number of already existing software components. These are coordinated and developed in a number of projects led by PSNC and the Polish national data storage (nds.psnc.pl).

Digital Library of Wielkopolska

### Objectives

The de facto deployment consists of the use of dArceo, and dLibra as user facing services, and the PLATON Popular Archive Service as a data storage solution. While this deployment configuration is known to work in a productive setting it is not known whether it will, without further development, satisfy the requirements of the DCH preservation roadmap coming out of WP3.

Alternatives exist in the CH community as well as in the Grid and Cloud e-Infrastructure communities that may be suitable as replacements for some or even all of the current components.

For example, the eCultural Science Gateway (eCSG) service (developed in the INVENT project) is very popular in other European member states such as Italy, Belgium and Sweden. For due diligence purposes this user portal component should be evaluated as part of the proof of concept.

On the data storage backend, several solutions exist. While the current PLATON service “Popular Archive Service” is known to work in production environments, there may be alternatives that satisfy the DCH preservation roadmap better and may have better sustainability options than existing solutions deployed in Poland. Next to the obvious PLATON deployment, large European Grid e-Infrastructures offer reliable distributed data storage services with a wide range of data retention policies. Similar offers are emerging in the Cloud Computing domain, where Cloud storage may offer advantages over Grid offers (or in fact have yet unknown disadvantages).

This Proof of Concept aims to test various combinations of deployments, combining existing Polish solutions (dArceo, dLibra, PLATON PAS) with existing and emerging alternatives (e.g. Grid or Cloud storage, eCSG user portal).

The main overarching goal of this Proof of Concept is therefore two-fold:

* Assess the suitability of existing Polish solutions compared to the DCH preservation roadmap,
* Examine alternative solutions and any necessary integration work

## Hungary (NIIFI)

# Conclusion

# Annex A – SCRUM: Agile Project Management

**Tables:**

|  |  |
| --- | --- |
| **Column 1** | **Column 2** |
| Text | Text |
| Text | Text |

Table 3: Sample table

1. http://en.wikipedia.org/wiki/Waterfall\_model [↑](#footnote-ref-1)
2. http://en.wikipedia.org/wiki/Planning\_poker [↑](#footnote-ref-2)
3. http://en.wikipedia.org/wiki/User\_story [↑](#footnote-ref-3)