**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 |
| Location | <https://documents.egi.eu/document/1544>  |

|  |  |
| --- | --- |
| Author(s) | Michel Drescher (EGI.eu) |
| Contributor(s) | Giovanni Ciccaglioni (ICCU), Sanja Halling (RA), Rosette Vandenbrouke (BELSPO), Indrek Eensaar (EVKM), Maciej Brzezniak (PSNC), Lajos Balint (NIIFI) |
| 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 | 21 Jan ‘13 | Michel Drescher (EGI.eu) | Integrated BELSPO contribution, integrated RA contribution, further process in section 3 |
| 0.6 | 23 Jan ‘13 | Michel Drescher (EGI.eu) | Added sections 3.4, 3.5, started section 3.6 |
| 0.7 | 24 Jan ‘13 | Michel Drescher (EGI.eu) | First pass of contributions to section 4 |

Table of Contents

1 EXECUTIVE SUMMARY 5

2 Introduction 6

2.1 Objectives of the deliverable 6

2.2 Structure of the document 7

3 Methodology, Tools and Processes 8

3.1 Audience and stakeholders 8

3.2 Agile Project Management vs Waterfall model 9

3.3 Timetable for activities 10

3.3.1 Contractual work packages deliverables and milestones 10

3.4 Roles and actors 11

3.4.1 Product owner 11

3.4.2 Facilitator 12

3.4.3 Project team 12

3.4.4 Project sponsor 12

3.5 Agile Toolkit 12

3.5.1 User Story 13

3.5.2 Epics 13

3.5.3 Tasks 13

3.5.4 Backlog 14

3.5.5 Sprints and sprint planning 14

3.6 Collaboration tools 16

3.6.1 Single sign-on 17

3.6.2 Mailing lists 17

3.6.3 Wiki 17

3.6.4 Document database 17

3.6.5 Meeting planner 17

3.6.6 Discussion forum 17

3.6.7 Blog 17

3.7 Iterations and Continuous testing 18

4 Proofs of Concept (PoC) 19

4.1 Italy (INFN, ICCU) 19

4.2 Sweden (RA) 19

4.2.1 Pilot lead partner 19

4.2.2 Environment and constraints of the pilot 19

4.2.3 Objectives 20

4.3 Belgium (BELSPO) 21

4.3.1 Pilot lead partner 21

1.1.1 Environment and constraints of the pilot 21

1.1.2 Objectives 21

4.4 Estonia (EVKM) 21

4.5 Poland (PSNC) 21

4.5.1 Pilot lead partner 21

1.1.3 Environment and constraints of the pilot 21

1.1.4 Objectives 22

4.6 Hungary (NIIFI) 22

5 Common assessment criteria 23

6 Conclusion 24

7 Annex A – SCRUM: Agile Project Management 25

# 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 : 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. Table 2 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 : Chronological planning of WP5 activities

As illustrated, the activity planning and coordination will take place over almost the complete project duration: In agile project management, planning continuously aligns current activities with the goals of the project that may (or may not) change. Frequent iterations and re-aligning with a more complete knowledge of the Proofs of Concept (i.e. the “problem space”) must accompany the actual work carried out by the partners in WP5 conducting the Proofs of Concept

## Roles and actors

In agile-managed projects, three key roles are working together to reach the objectives of the project. Together they *steer*, *facilitate* and *execute* the necessary work until the end of the project.

### Product owner

In pure agile projects the *product owner* is a single key person equipped with the authority to steer the direction of the project and its outcomes. The product owner does so by ensuring a constant supply of background material and user stories coming from the target domain of the product. The product owner further prioritises user stories and thus steers, which topics and tasks in the project shall be, tackled first and which later. Finally, the product owner is empowered to accept or reject a delivered solution to a task in the sprint review meeting (see below) and must be prepared to report on this to the users and customers she is representing in the project.

While the product owner is often and by design and intent one single person, it is at times practical to accept two, rarely three, product owners in a project. However, if done so that instigates the obligation on all appointed product owners to commit to intensive communication with each other and to speak with one voice towards the other roles in the project.

### Facilitator

The *facilitator*, in the agile SCRUM methodology also called SCRUM Master, fulfills in agile projects the role of a project manager in classic project management. The key objective of the facilitator is to enable the product owner and team (team members, really) to work as effectively as possible. The facilitator is expected to do so by providing the tools, processes, techniques and reports on the project, but *not* the content. This is the duty of the product owner (see above) and, most significantly, the team.

The Facilitator calls for and chairs meetings, plans sprints and communicates intensively with the product owner and the team. The facilitator also regularly reports to the *project sponsor* (see below). One key activity of the facilitator is to maintain the project backlog as accurate as possible, and both the team and the product owner are obligated to provide the necessary information as requested by the facilitator.

### Project team

The project team comprises of individuals carrying out the work that is captured in the project’s backlog, tasks and sprints. It is desirable that the skill sets of the team members are diverse, but not disparate, with sufficient overlap. Otherwise, the project bears the risk of domain specialists incapable of understanding each other, thus impeding the common commitment and progress of the whole project.

The team is considered to act as one unit, and is expected to organize itself over the course of the project. The product owner(s) and the facilitator are always available for the team to resolve issues and problems for them.

### Project sponsor

Often overlooked, the project sponsor is the entity that provides the financial backing and strategic authority to grant the project to take place. It is the authority involved in team recruiting & management, risk management and financial management of the project. In case of disputes the project sponsor is the arbitrating instance between product owner, facilitator and the team, and if necessary the project sponsor is the entity may decide to shut down a project at any time.

In smaller projects, it is often the same (group of) people exercising the roles of project sponsor and product owner even though they provide distinct value. Larger projects are often hierarchically organized as agile projects of agile sub-projects (in the SCRUM methodology called “SCRUM of SCRUMS”), where the parent project acts as project sponsor to all its sub-projects.

## Agile Toolkit

Key fundamental principles of agile project management methodologies are *dedication, communication* and *collaboration*. These three social functions enable the team members to achieve the goals and objectives the team aspires to reach. All techniques and tools employed in agile methodologies are used as *facilitators,* i.e. supporting and encouraging communication and collaboration – which cannot happen without *information*. Thus, information needs to be provided and maintained, and made publically available for anyone interested.

Not all agile tools and techniques are applicable to this project and work package. Common agile management seeks to bring the agile team physically close together to maximize team dynamics and effectiveness. This physical proximity allows employing certain agile techniques, such as daily stand-ups daily peer-review of work. Albeit not impossible it will be difficult to introduce such activities in a distributed project setup such as DCH-RP. After careful consideration the following agile techniques and tools will be used to conduct the work in Work Package 5. They are not documented in full detail; they are captured in this deliverable with respect to the intent of the tool and the applicability to this project.

### User Story

A User Story captures a metaphor of value of the system to the end user. User Stories are not intended to provide full documentation, they should remind the team members of what needs to be captured and taken care of. In essence, User Stories should be seen as the start of a *discussion* between team members, forming a commonly shared mindset across the team about what the outcome is supposed to deliver.

User Stories are related to Use Case descriptions; User Stories are by intent kept in the narrative language of the target domain (i.e. in this project, the language and terminology of the DCH community) albeit somewhat informal in their content. Use Cases, on the other hand, provide more formal information on the interaction of the user with the system – within the scope of this project that would be the DCH institute managers and colleagues who use the system to *preserve* and *publicise* DCH data for other users to work with.

User Stories will have to provide three aspects: A title, the description, and a “definition of done”. A number of guidelines exist on how to write good user stories, and these will be provided to project partners elsewhere. However, the title captures the essence of the user story, the description establishes the metaphor capturing the aspired customer value, whereas the third section (the “definition of done”) essentially provides a description of the acceptance criteria and/or indicates test descriptions, answering the question “How and when will we know that this user story is satisfied?”

User stories are continuously provided and refined by the product owner of the project.

### Epics

In large, complex projects, it can become difficult to find the right metaphor to capture customer value purely in User Stories. A common activity in agile projects is breaking down User Stories in better-scoped sets of User Stories. With the number of User Stories growing, patterns and clusters of related User Stories begin to emerge. To keep an overview of the knowledge space, such User Story clusters are grouped into “Epics”, each describing the overarching goal.

Both User Stories and Epics are subject to prioritisation by the consuming stakeholders – in the DCH-RP project that translates to WP3 and WP4 – influencing and steering how WP5 will conduct its work.

Epics – if applicable – are often a collaborative work of the product owner and the team.

### Tasks

Tasks are central work planning artifacts and describe what needs to be done in order to satisfy the supplied user stories. Often, User Stories are turned directly into Tasks. During the course of a project tasks emerge that capture routine work, at times called chores that are not directly linked to user stories. Nonetheless these routine tasks need to be discussed in the regular sprint planning meetings, and tracked throughout the sprint execution.

Central to agile management is a concept of capturing the necessary effort to accomplish the task in an abstract manner called *story points*. The more story points are assigned to a task, the more effort is deemed necessary to complete it. A number of story point schemes are popular and neither is superior to others. The two post popular schemes are a linear scheme, and a Fibonacci scheme. The linear scheme is very easy and assigns story points based on the linear scale from 0 to 10. The Fibonacci scheme (or derivatives) uses the Fibonacci numbers (including 0) up to 13 (i.e. 0, 1, 2, 3, 5, 8, 13). Fibonacci scheme variations then often utilize the numbers 20, 40, 60, 100 as risk management flags (see below).

Tasks are the entire responsibility of the product team alone; by adding to or removing tasks from the backlog, and adding to and removing tasks from sprints the team takes responsibility and commits to the progress of the project.

### Backlog

During the course of the agile managed project the team keeps track of all tasks in an artifact called *project backlog*. This project backlog is pivotal to the success of the managed project and captures the following concepts:

* Project scope and complexity
	+ Complete picture of all user stories and epics
	+ Complete documentation of all tasks
	+ Prioritisation
* Historic information
	+ Completed tasks
	+ Past sprints
* Project management
	+ Current project velocity and team capacity
	+ Project duration projection (velocity to remaining story points ratio)
* Risk management
	+ Which tasks are not prioritized
	+ Which task are flagged as insufficient (using story point risk markers)
	+ Prioritisation to keep project duration projection in scope with hard deadline

The information presented in the backlog is kept at a minimum but facilitates exercising the above-mentioned concepts based on that information given in it. It is often maintained as an overview document providing pointers and references to detailed information stored elsewhere. This document/artifact is updated very frequently, reflecting the project’s agile activities. It must be kept as accurate as possible.

The backlog is the key tool used and maintained by the project facilitator.

### Sprints and sprint planning

A sprint in agile project management captures a single unit of organized collaborative work. An agile project as a whole is organized as a series of consecutive sprints in a forward-chain towards the end of the project. Agile-managed projects typically end either when all tasks are completed (value-oriented open ended projects), or with a deadline defined and agreed prior to the start of the project (time-constrained projects). Obviously, the DCH-RP project is a time-constrained project, so risk-management and prioritisation become more important than feature inclusion.

Sprints have a fixed structure in agile project management and it is of paramount importance that this structure is kept at all times. Otherwise, the project is in danger of wasting excessive effort on structure and management rather content. Even though an agile project would not be agile if the sprint structure would not be eligible to change, changes should nonetheless be introduced carefully and gradually in order to maximize the team’s effort and output (as measured by its velocity).

Sprints shall be planned to identical length. Although deviations are common they must be kept to a minimum, reducing impact on the team rhythm and velocity calculation. In each sprint, the following phases and meetings take place, in the following order:

1. **Sprint Planning**

This meeting involves all team members. In this meeting, team members discuss tasks listed in the project backlog that are prioritized. The team discusses unestimated tasks one by one until it reaches a common understanding on the necessary effort. It is the team’s responsibility to capture discussion outcomes wherever and however it deems appropriate for further reference. Once agreed, the team estimates the effort by assigning story points to the respective task. If the team cannot agree then a risk flagging estimate will be assigned (i.e. estimates greater or equal to 20) for further assessment. The goal is to provide the team with enough work to do for the planned sprint – but not more. This is done by including as many estimated tasks into the sprint so that the sum of story points does not exceed the current team velocity. Usually, the team will discuss tasks according to priority, and picks those that will be tackled in the current sprint. It is important that the whole team agrees, as the whole team will commit to this work plan for the current sprint.

The outcome will be recorded in the backlog (i.e. which tasks will be tackled in which sprint) and the meeting will close.

Sprint planning meetings should not take more than a couple of hours (half a day at max), albeit teams new to agile project planning tend to need more time in the beginning.

Participants of a spring planning meetings are the Facilitator and the complete Project Team. The Product Owner is optional, but highly recommended to participate. Observers are allowed to attend, but not permitted to interact with or influence the participants. The Facilitator is empowered to exclude observers from the meetings, whether temporarily or permanently.

1. **Sprint Execution** (i.e. working on tasks)

This phase comprises the majority of the time in the sprint. During that time the team is working on the tasks agreed upon in the sprint planning meeting.

*The team will work in a self-organising manner and as independent and uninterrupted as possible during this phase.*

Tasks will be worked on, and marked as complete by the team members. As a general rule, every task outcome will be reviewed by another team member; when rigorously followed this process ensures not only built-in quality assurance but also facilitates the common and complete understanding of the project within the team.

Tasks are usually not assigned to team members; instead Team Members proactively take tasks and mark their name on them to indicate that they are working on them. When they are done, they mark them as *Delivered*, waiting for the peer review. Once that is done,

1. **Sprint Review**

At the end of every sprint, the Facilitator calls for a sprint review where the Project Team reports and presents to the Product Owner what has been achieved in the ending sprint. Participants also discuss what was *not* achieved, problems and missing information related to the project. The Product Owner determines whether the work pertaining to a task or user story is accepted or not. The Facilitator records these decisions in the affected tasks and user stories. Project Team and Product Owner are responsible for taking notes by themselves; meeting minutes are considered not providing value to the project in agile project management.

Tasks rejected by the Product Owner are put back into the Project Backlog and marked as not done. Tasks that are accepted by the Product owner are marked as finished. When all tasks that are part of the current sprint were reviewed, the story points of all finished tasks are summed up to the *sprint performance*. The last step in the Sprint Review is the calculation of the *Team Velocity* as the average over the sprint performances of the last three sprints. This updated velocity will then be recorded and used for planning the upcoming sprint.

If time permits, the Project Team and the Product Owner can further work on User Stories and Tasks, Prioritisation, Task estimation and any project related topics.

Sprint reviews should not last longer than half a day; if team velocity and achieved tasks indicate that more time might be needed the meeting material should be prepared to keep the half-day limit.

The participants of the Sprint review are the Product Owner, Facilitator and the Project Team; attendance is compulsory. Again, observers are allowed with identical restrictions as described above.

1. **Sprint Retrospection**

The Sprint Retrospection is the only optional component of a sprint. Albeit allowed, it is strongly discouraged to skip this phase of the sprint.

As with all other meetings the Facilitator chairs the Sprint Retrospection meeting; attendance is compulsory for the Project Team. This is the only closed meeting where no other participants and observers are allowed.

The objective of the Sprint Retrospection is to continuously optimize group dynamics, communication, processes, and methodology; anything that helps improving the team’s overall sprint velocity. Nothing is out of scope when the team believes that a change will contribute to their overall satisfaction and happiness (which in turn have a positive effect on the team velocity). Guided by the Facilitator, the Project Team Members examine the passed sprint and identify “the good, the bad, and the ugly”. Things that went (perhaps horribly) wrong, and things that went (hopefully very) well, are discussed in an open and encouraging atmosphere (for which the Facilitator is responsible). From that discussion, the Project Team agrees on a number of changes that each and every team member will commit to.

These outcomes are recorded[[4]](#footnote-4), but kept from wider circulation – it is the Product Teams own material to organize itself, not for others to interfere with. The Facilitator, however, is allowed to use that material for subsequent retrospection sessions to be able to guide the meeting and bring the team’s attention to patterns of issues that otherwise might be undetected.

Therefore the material is kept separate from other, otherwise public information. In the (rare) event of intervention through the Project Sponsor, the Product Team may decide to disclose the material coming out of the Spring Retrospection meeting.

## Collaboration tools

Work Package 5 will make use of a number of collaborative tools that are provided by EGI.eu as part of their contribution to the DCH-RP project[[5]](#footnote-5). These have been proven very useful to EGI.eu and the EGI-InSPIRE project as tools to facilitate collaboration in project management, documentation, and coordination. After consideration the DCH-RP project management decided to take that offer and make these tools available for all project members.

The DCH-RP project is pursuing two objectives with this integration:

1. Provide the project work packages and members with a powerful set of collaboration tools, and
2. Evaluate the available tools for inclusion in the preservation roadmap, which is the key output of Work Package 3.

The following subsections briefly illustrate each tool, and what it is used for within the DCH-RP project.

### Single sign-on (SSO)

This tool provides the central integration point of all other offered tools. It is a facility providing login credentials to potentially unlimited numbers of individuals. Next to identity management this tool provides user groups, managed by group owners. Users can be added and removed to any number of groups. Identity is fundamentally linked to valid Email addresses. Each user can freely choose a username for as long as it is unique at creation time. This has the consequence that a user has to register and deregister when his or her Email address changes. Although this may happen, it usually happens infrequently, and the transition to a new Email address is typically fully accomplished fairly swiftly.

The key integration point with other tools is the groups and group membership information provided by this service.

The EGI SSO service is accessible at [https://www.egi.eu/sso](http://www.egi.eu/sso) for anyone with an SSO account. SSO group owners manage SSO groups and membership through the same address.

### Mailing lists

For each SSO group exactly one mailing list may be enabled. All mailing lists are operated using the Open Source tool “mailman” and inherit their names from the SSO group, under the common Mailing List server “mailman.egi.eu”. Consequently, the mailing list for the SSO group “dchrp-all” is “dchrp-all@mailman.egi.eu”. All mailing lists offer the feature-rich standard mailman administrative interface.

Mailing list membership is managed through SSO group membership; when adding an individual to a specific SSO group, and a mailing list for this SSO group is enabled, that individual will be automatically subscribed to that mailing list. Conversely, if an individual is removed from an SSO group, mailing list membership will be revoked, too. This functionality is the same for all integrated collaborative tools provided by EGI.eu.

Currently it is not clear whether the DCH-RP project will make extensive use of mailing lists (and SSO groups). It is currently foreseen to provide a general mailing list and one mailing list per DCH-RP work package. This pattern has proven useful for the EGI-InSPIRE project to limit the amount of Emails people receive that are not within their scope of work.

With respect to WP5 a scoped SSO group and mailing list will be implemented and used as general Email based coordination across the work package.

### Wiki

Using the popular Mediawiki software EGI.eu offers Wiki functionality to the DCH-RP project. As with all other integrated tools, the Wiki user accounts are managed through the EGI SSO functionality. For the DCH-RP project a separate namespace (“DCH-RP:”) was implemented. While read access to this Wiki namespace is public, all members of the “dchrp-all” SSO group automatically have write access to this namespace (but not to the default, or any other namespace). While the DCH-RP project shares the same Wiki service instance with all other customers, the project has exclusive change control over all Wiki entries within its own namespace.

Work Package 5 will use the Wiki as a collaborative workspace for any kind of documentation with high to medium update frequency. We expect to store and maintain user stories, tasks and epics in the Wiki, as well as any feedback and rationale provided by product owners for accepting or rejecting tasks during the sprint reviews. Additional task related information, such as sub-tasks (if any), notes, etc. will be stored together with the notes.

The DCH-RP specific Wiki is accessible at the address [https://wiki.egi.eu/wiki/DCH-RP:Main\_Page](https://wiki.egi.eu/wiki/DCH-RP%3AMain_Page)[[6]](#footnote-6).

### Document database

The primary purpose of the Document database (DocDB for short) is to provide reliable, permanent storage locations for documents that are edited and maintained offline but need public circulation and reference. EGI.eu provides this service through the Open Source product “DocDB”. EGI uses DocDB extensively for all kinds of documents (text documents, presentations, spreadsheets, etc). DocDB does not limit the type of data stored for as long as it is a well-defined octet stream (i.e. with a beginning and an end). Albeit users usually manage one actual electronic document per DocDB document, any number of related documents may be grouped and maintained in one DocDB document[[7]](#footnote-7). Each DocDB document is accessible via a persistent public address; for example this deliverable is accessible using the address or location <https://documents.egi.eu/document/1544>. DocDB documents are identified using a persistent numeric identifier (e.g. 1544). Extensive Metadata allows tagging DocDB documents according to the user’s specific needs.

The DCH-RP project intends to use DocDB for storing and sharing documents such as deliverables, offline documentation, and presentations. For this a number of DocDB topics were created for the DCH-RP project[[8]](#footnote-8).

### Meeting planner

### Discussion forum

### Blog

The major contributions to Work Package 5 will originate from Work Package 3 and Work Package 4 as necessary material for Work Package 5 to digest into actionable tasks. While the bulk of these contributions will be provided as background material that is expected to evolve in an iterative manner, WP3 and WP4 are expected to capture that material in metaphors described in User Stories and Epics (see below) as ancillary material.

Agile activity management artefacts

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

## Iterations and Continuous testing

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)

To prepare the work for the Proofs of Concepts, we need to acquire an initial understanding of each partner’s role in their local context, and which expertise and skillset is available in each partner. As this is a distributed effort across Europe, we expect cultural and socio-economic differences among partners and their local data providing institutes, be it in the target domain of cultural heritage (type of heritage, archiving and preservation methodologies, digital data formats etc.), conduct of work and available e-Infrastructure.

Hence distinct objectives among the project partners are likely, although sufficient overlap in common goals is expected.

The following subsections illustrate each partner and partner network, and provide summaries of the objectives of each partner. These sections are designed to be individual, though following a common structure for better comparability. Section 5 will distill these individual presentations and identify a starting set of common criteria and objectives among all partners.

## Italy (INFN, ICCU)

### Pilot lead partner

The Central Institute for the Union Catalogue for the Italian Libraries and for Bibliographic Information (ICCU) promotes and coordinates cataloguing ad documentation activities of the library heritage under the authority of the Ministry for Cultural Heritage and Activities. It coordinates national and international projects, one of which, Internet Culturale, is involved in this project, with its staff and digital assets – <http://www.internetculturale.it>.

Internet Culturale is the Italian digital library and cultural institutions portal. It provides “digital and multimedia resources” to delve into literary, scientific, artistic, and musical culture. Internet Culturale makes it possible to view, in the same “virtual site”, catalogue information and digital collections produced by its partners.

### Context and constraints of the pilot

Cultural institutions can become partners of Internet Culturale and have their own digital archive. In this case metadata are indexed by Internet Culturale’s services through the standard protocol OAI-PMH[[9]](#footnote-9). The digital objects that reside in the peripheral archives are pulled up each time the viewer is activated. Partner institutions can also take advantage of a service made available by Internet Culturale: the MagTeca. This digital archive provides a free management and conservation service for digital collections at web resolution and complete with metadata encoded in xml under MAG standards.

We expect the system to follow as much as possible the larger number of the objectives described below, but above all we expect that it could provide the best tools and resources available, in order to assure the chain of preservation of our digital assets.

To sum up, the pillars on which the system should be developed are:

* security;
* fastness;
* migration to new formats;
* documentation

### Objectives

For our digital archive, the Mag Teca, we need to store and preserve the digital assets, i.e. metadata plus web copies of digital objects, produced by our partner institutions. We want to be always ready to make changes and updates due to upgrades of technologies.

The goal is to offer the best service to our partners and, above all, to our patrons and end users.

Therefore, we expect that the system, i.e. e-infrastructures, staff members and economic resources, could provide:

* A fast and secure transmission of the digital assets during the upload and the download (priority level 9/10);
* A data integrity checking tool, to provide the assurance that the files (data plus metadata) have arrived intact during the upload and the download (priority level 9/10);
* A staff user-friendly GUI, with multiple fields for data entry and retrieval (e.g. name of collections and agency, topic/subject, type/format, both of digital objects and metadata, amount etc.) and multiple steps (e.g. a review of data prior to uploading and to downloading) (priority level 7/10);
* Scheduled reports (e.g. monthly, bimonthly, etc.) on the status of the digital assets (priority level 6/10);
* The management of preservation metadata, in order to assure constant maintenance and migration to new formats and technologies (priority level 9/10);
* A system of different, multilevel staff privileges, from a top level for the collection manager/records keeper, down to the other staff member (priority level 7/10);
* A warning system that alerts the collection manager/records keeper whenever the digital assets are modified, downloaded etc. (priority level 7/10);
* A clear policies document to be followed by the organization during the workflow (priority level 8/10);
* Documentation activities with the history of all actions performed on the resources, including changes and decisions (priority level 9/10);
* A staff always up to date, and expert with the needs of cultural heritage institutions (priority level 6/10)

## 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.

### Context 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.

### Context 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)

### Pilot lead partner

The Estonian Ministry of Culture (EVKM) is the project partner responsible for the organisation of Proof of Concepts in WP5. EVKM is not a cultural heritage organization itself and neither does it manage collections. Conservation Centre KANUT has shown interest in participating in the pilot action. KANUT carries out digitization work for different museums and engages in 3D digitization.

### Context and constraints of the pilot

The Conservation Centre KANUT has a lot of experience in archiving digital data in cooperation with the Estonian Public Broadcasting. Conservation Centre KANUT uses services provided by the Estonian Research Network (EENet) but has not yet the experience of using EENet e-infrastructure for data archiving.

### Objectives

The key issue for the Conservation Centre KANUT is to test different data, including that of 3D models, so that on a positive result of PoC the basis is laid for a sustainable DCH data infrastructure and corresponding services. Conservation Centre KANUT is also interested in long-term cooperation with the EENet in order to solve the issue of long-term data preservation. Conservation Centre KANUT is also interested in cooperation with the EENet regarding 3D modelling. Cooperation concerning 3D models for data storage in EENet would be an excellent starting point for continuous cooperation.

The common definition of tools and services should also accelerate an even better exploitation of the archived data.

## Poland (PSNC)

### Pilot lead partner

**PSNC** (Poznan Supercomputing and Networking Centre, affiliated to Institute of Bioorganic Chemistry) has multiple roles in Polish environment. PSNC is NREN (National Research and Education Network) developer and operator, HPC (High Performance Computing) center, R&D organization and services provider. PSNC provides network, computing, storage, grid and cloud facilities to the academics community in Poland. PSNC also develops and supports software for data and content management.

PSNC manages the PIONIER[[10]](#footnote-10) network that spans 700+ R&D institutions and universities delivering them 10Gbits interconnects and 40GB/s uplinks to European Internet hubs and international links to Czech Republic, Lithuania, Sweden and other countries, through 7000+km of optical fiber infrastructure.

PSNC participates in national and international projects related to data and content management, grids and clouds, including EGI-Inspire, PL-Grid, EUDAT, National Data Storage[[11]](#footnote-11), PLATON[[12]](#footnote-12) as well as EuropeanaCloud[[13]](#footnote-13), LoCloud and SYNAT[[14]](#footnote-14), IMPACT[[15]](#footnote-15), SUCCEED and Wf4Ever[[16]](#footnote-16).

Digital Library of Wielopolska[[17]](#footnote-17) (**DLoW**) is the largest digital library federation in Poland. Currently it holds 300 000+ of publications of different types including cultural heritage objects such as manuscripts, rare books, Polish historical documents, cartographical collections as well as educational materials including schoolbooks, lecture texts, monographs and regional materials related to Wielkopolska, e.g. leaflets, guides, posters, catalogs of exhibitions and fairs etc. DLoW is the joint initiative of the Poznan academic community, formally managed by Poznan Foundation of Scientific Libraries[[18]](#footnote-18) (PFSL). PSNC has very good contacts with DLoW and PFSL members and will exploit them for the DCH-RP PoC purposes.

### Context and constraints of the pilot

The Polish CH community currently uses a number of already existing, mature data management and content handling tools, services and systems. They are coordinated and developed in a number of projects led by PSNC and its partners from PIONIER network consortium.

Within EGI-Inspire, PSNC provides 400+TB of disk storage space for gLite and other VOs. For instance DCH-RP VO is entitled to use these storage resources for the PoC purposes.

Under PLATON project, PSNC runs **Popular Archive Service** with capability of 12+ PB of tape and 2+ PB of disk storage and network file system like-interfaces (sFTP, WebDAV, GridFTP). This service is run on top of dedicated infrastructure including storage (10 disk arrays, 5 tape libraries), computing (80+ servers) and network (20+ 10Gbit and FC switches), spread across the country. Data access, transparent replication, integrity checks as well as reliability, security and data persistency mechanisms are implemented in National Data Storage software, developed and improved by PSNC and partners since 2007. These services and solutions provide reliable, replicated and safe storage of the files, directory structures and simple user-level meta-data (e.g. annotations to files).

PSNC develops **dLibra** (Digital Library Framework) another state-of-the-art solution for content management and delivery. dLibra allows building repositories of digital documents which can be accessed by external individuals and systems on the Internet. dLibra is de-facto standard software in Poland, as it is run in 100+ instances and used by hundreds institutions in Poland, including cultural heritage institutions. Notably, 98% of the total 1,1M of the digital objects served by Polish institutions is delivered via services based on dLibra software.

Following the dLibra concept, PSNC develops **dArceo**, **dMuseion** and **dLab** products, which deal with long-term data preservation and domain-specific issues of data and content management. In particular, dArceo, is an OAIS-compliant solution for long-term preservation of source data, supporting meta-data extraction, data format migration and conversion process management, integrity control (including SHA-512 checks, format verification using UDFR, i.e. PRONOM and GDFR), advanced data delivery, including optimization of presentation versions for particular end-user devices, etc.

Importantly, dArceo, supports various storage back-ends, including sFTP. This enables PSNC and partners to deliver a multi-level solution for long-term data preservation, combined out of dArceo (on the content management level) and Popular Archive Service (on the data persistency, safety and storage reliability level). We predict that most of the institutions currently using dLibra for content delivery purposes will also use dArceo and Popular Archive Services for protecting their Archive information Packages and managing the long-term preservation processes.

Digital Library of Wielkopolska (DLoW) will be partnering institution of PSNC in the DCH-RP PoC. The PoC will use part of the more than 300 000 objects managed by this library. Types of the objects that may be made available by DLoW to our pilot include: general collections such as books, regional periodicals etc. as well as special collections including manuscripts, notes, maps, music, and photography albums. Type of the material used for PoC purposes and scope of the collections used for PoC will be agreed with DLoW according to project needs identified on the further stages of the work.

### Objectives

The de facto standard tools used by CH community in Poland consist of dLibra as the user interface, dArceo as the long-term preservation toolset and the PLATON Popular Archive Service as a long-term, reliable data storage service. This deployment configuration is known to work in a production setting. It also addresses the typical requirements of the cultural heritage data preservation process. Notably, these solutions were and are being designed and developed by PSNC and its partners in the close collaboration with DC community members.

Therefore, the practical analysis of the functionality and features of the solutions existing and developed in Poland versus the DCH preservation roadmap coming out of WP3 of our project, will lead to two kinds of conclusions. First, preservation roadmap elements not yet addressed by existing products will be analyzed in the confines of PoC. Missing elements and functionality gaps will be considered for inclusion in the developments plan of these tools. Second, functionality, features, approaches and solutions worked out for Polish CH community can be fed up to the DCH preservation roadmap.

Within WP5’s PoC we will perform functionality, scalability, sustainability, interoperability and user-friendliness oriented comparison study of eCultural Science Gateway (eCSG) service (developed in the INVENT project) and dLibra/dArceo and PLATON Popular Archive Service as well as the National Data Storage system concept. We will involve both dLibra/dArcheo and PLATON developers and the system end-users in this process. This work may, again, lead to bi-directional feedback regarding the eCSG and dLibra/dArche/PLATON/NDS features and characteristics. We will also collect the knowledge and experience related to desirable vs realistic, sustainable service deployment and offering models (e.g. cloud model vs Grid model, private IaaS/SaaS cloud vs public SaaS) as well as usage of standards and techniques appropriate to data preservation aims (e.g. data access protocols typical to Grids vs clouds).

Part of this study will be to examine, how the user front-ends and content management-related modules of eCSG, dLibra and dArceo can be combined with and take benefits of existing and future Polish and European data storage and management services and infrastructures such as PLATON Popular Archive Service and EGI-Inspire/NGI-managed Storage Elements. Cloud Storage systems developed by PSNC, other European NRENs, EGI and other initiatives (such as e.g. EUDAT) will be also taken into account.

At the data storage end we will analyze distributed storage system mentioned above from the point of view of long-term data preservation supporting features such as data replication and possible data retention policies, data integrity control, versioning, disaster recovery etc. We will also analyze the various approaches for assuring the sustainability of the data storage platform, storage resource provide-independence and avoiding vendor lock-in and storage provider lock-in – issues which are critical from the point of view of long-term data storage and preservation.

Overall this Proof of Concept aims to test various combinations of deployments, combining existing Polish solutions (dArceo, dLibra) with existing and emerging alternatives and tools (e.g. Grid or Cloud storage, eCSG user portal, PLATON PAS etc.). Details of particular PoC scenarios will be agreed with CH institution and DCH-RP project partners at the later stages of the work.

## Hungary (NIIFI)

### Pilot lead partner

NIIFI (National Information Infrastructure Development Institute) is the Hungarian participant at the DCH-RP project, responsible for providing the network, processing and storage infrastructure in Hungary for DCH preservation activities while OSZK (Országos Széchényi Könyvtár – National Széchényi Library) is the Hungarian partner of NIIFI in DCH-RP, as major representative in the country of performing DCH collection and preservation.

NIIFI is the NREN (National Research and Education Networking organisation) in Hungary and also serves as NGI of the Hungarian research and education community. NIIFI handles the access for the related community (research, education, public collections) to a wide range of national and international network services, operates HBONE+, the community’s countrywide private 10+ Gbps backbone network, and provides international connectivity (GÉANT) to the entire community. NIIFI offers a wide service portfolio including advanced videoconferencing, VoIP, grid/ClusterGrid, HPC and storage cloud, digital libraries/archives services etc. IPv6 is widely used on the DWDM-based backbone and the access network using multiple technological variants.

OSZK is the largest collection of written heritage and related documents in Hungary. The mission and task of OSZK are to collect, process and preserve all the documents belonging to that category of cultural heritage, with special emphasis on DCH related to Hungary (books and other documents published in Hungary, as well as publications related to Hungary or to Hungarians published abroad). The collection is available for the users whether in the form of parchment, paper or electronic medium.

### Context and constraints of the pilot

NIIFI provides the basic countrywide infrastructure with international connections. Major technical characteristics (environment, architecture and tools) of the infrastructure are summarised below:

* N x 10 Gbps hybrid network (IP + e2e)
* 3200 km DF (CEF) backbone (→ GÉANT)
* N x 10 Gbps production, 40-100 Gbps experimental traffic
* 79 PoPs (+ GÉANT PoP Budapest)
* About 500 institutions (~ 700.000 users), to be drastically increased (schools) from 2013
* Complex AA (authentication and authorisation) infrastructure
* Grids (including first ClusterGrid in Europe
* Cloud services (non-commercial)
* ~ 50 Tflops aggregate HPC
* ~ 2.5 PByte distributed storage
* Advanced collaboration facilities (~ 130 HD VC nodes + desktop option)
* Traditional Data Infrastructure relations (hundreds of public collections)

Major NIIFI cloud features for various application areas (including archives) are listed here:

* NIIF IaaS Cloud = distributed resource mgmt + virtualised machines + network;
* Private cloud for NREN users
* OpenNebula + KVM - open source
* Web mgmt interface
* 45 physical nodes – VPLS network
* ISCSI storage

OSZK is the DCH institution devoted to collection, processing, and preservation, as well as service provision to make the collection items available for the users.

Features of OSZK collections/archives are as follows:

* Stores own content (primarily library items)
* Builds both dark and operational archives
* Connected to NIIFI (e-Infrastructure provider)
* Active AAI user
* Specific mission and function – relevance for testing

Data types covered by the OSZK collections are, among others:

* General collections
	+ Books
	+ Periodicals
* Special collections
	+ manuscripts
	+ small prints
	+ microfilms
	+ early books
	+ theatre history collection
	+ map collection
	+ music collection
	+ historical interviews
* Digital collections – OAI-conformant repositories (OAI-PMH conformance)
	+ Hungarian Electronic Library
	+ Periodicals Data Base
	+ Multimedia Collection

### Objectives

The PoC investigations are supposed to cover (on top of the aspects and scientific domains listed in the Introduction) primarily the following aspects (specific PoC objectives):

* Sustainability:

Assumable usability, life-time, and cost/performance of applied methods, tools, protocols, standards, architectures (physical and virtual), etc.

* Interoperability:

Experiences with respect to joint exploitation of the OSZK and the DCH-RP environments (national and international infrastructure, preservation tools, storage options, etc.)

* User-friendliness:

Ease of use, interaction convenience, data accessibility, system reliability and dependability, federative AAI opportunities (SSO), coverage, etc.

Preservation opportunities and options are to be investigated in view of short-, medium- and long range applicability, by taking into account foreseeable evolutionary and possible revolutionary changes in the philosophical, technical, and practical approaches applied to preservation processes, projects, and programs. Special emphasis is to be put on international aspects (coordination, co-operation, collaboration, joint efforts, common goals, etc.)

Selection of covered collection types to investigate, and selection of well-defined, concrete objectives, well applicable and promising tools, as well as data to be used, will be made during the next phases of the DCH-RP project.

# Common assessment criteria

- initial list of common criteria to assess results

 --> add to D5.1

 --> usability of technical tools

 --> service feasibility (coming from D3.1

 --> preservation tasks, preservation metadata, obsolete ("dark") data

 --> data integrity, etc.

# Conclusion

The conclusion.

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)
4. The Facilitator may provide templates and storage location for the material, but it is the Product Team’s responsibility to record the discussed topics. [↑](#footnote-ref-4)
5. For the duration of the DCH-RP project, EGI.eu will guarantee access free of charge to the collaboration tools, for all project members, on the basis of their current deployment. [↑](#footnote-ref-5)
6. Note the use of the “DCH-RP:” namespace in the name of the DCH-RP Wiki landing page. [↑](#footnote-ref-6)
7. For example, see <https://documents.egi.eu/document/390> [↑](#footnote-ref-7)
8. See also <https://documents.egi.eu/public/ListTopics> for details [↑](#footnote-ref-8)
9. Open Archive Initiative (OAI) Protocol for Metadata Harvesting (PMH) – see <http://www.openarchives.org/pmh/> [↑](#footnote-ref-9)
10. Polish Optical Internet. <http://www.pionier.net.pl/online/en/> [↑](#footnote-ref-10)
11. National Data Storage. <http://nds.psnc.pl> [↑](#footnote-ref-11)
12. PLATON - Science Services Platform. <http://www.platon.pionier.net.pl/online/?lang=en> [↑](#footnote-ref-12)
13. <http://europeana.eu> [↑](#footnote-ref-13)
14. <http://www.synat.pl> [↑](#footnote-ref-14)
15. IMPACT Competence Centre. <http://digitization.eu> [↑](#footnote-ref-15)
16. <http://wf4ever-project.org/web/guest/home> [↑](#footnote-ref-16)
17. <http://www.wbc.poznan.pl/dlibra?action=ChangeLanguageAction&language=en> [↑](#footnote-ref-17)
18. <http://www.pfsl.poznan.pl/en> [↑](#footnote-ref-18)