



Project Acronym:	DCH-RP
Project Title:	Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for DCH in 2020
Contract Number:	312274
Starting date:	1 st October 2012 Ending date: 30 th September 2014

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:	January 2013

Project Coordinator

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	31 st January 2013
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	Borje Justrell, & Raivo Ruusalepp (RA), Rosette Vandenbrouke (BELSPO), Lajos Balint (NIIFI), Roberto Barbera (INFN)
Approved by:	Antonella Fresa (PROMOTER)

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/RA contributions, further process in section 3
0.6	23 Jan '13	Michel Drescher (EGI.eu)	Progress on sections 3.4, 3.5, and 3.6
0.7	24 Jan '13	Michel Drescher (EGI.eu)	First pass of contributions to section 4
0.8	25 Jan '13	Michel Drescher (EGI.eu)	Final draft for review.
0.9	29 Jan '13	Michel Drescher (EGI.eu)	Interim reviewer comments integration
1.0	30 Jan '13	Michel Drescher (EGI.eu)	Integrated reviewer comments.
1.1	30 Jan '13	Claudio Prandoni (Promoter)	Final review and formal check.

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY.....	5
2 INTRODUCTION	6
2.1 OBJECTIVES OF THE DELIVERABLE.....	7
2.2 STRUCTURE OF THE DOCUMENT	7
3 METHODOLOGY, TOOLS AND PROCESSES	9
3.1 AUDIENCE AND STAKEHOLDERS	9
3.2 TIMETABLE FOR ACTIVITIES.....	10
3.3 AGILE PROJECT MANAGEMENT VS WATERFALL MODEL.....	12
3.4 IMPLEMENTING SCRUM IN WORK PACKAGE 5	13
4 STAKEHOLDER BACKGROUND AND OBJECTIVES	16
4.1 ITALY (INFN, ICCU).....	16
4.2 SWEDEN (RA)	18
4.3 BELGIUM (BELSPO).....	19
4.4 ESTONIA (EVKM)	20
4.5 POLAND (PSNC).....	21
4.6 HUNGARY (NIIFI)	24
5 PROOFS OF CONCEPT PLANNING OVERVIEW	27
5.1 PROOF OF CONCEPT EXECUTION PLANS	27
5.2 FEEDBACK TO PRESERVATION ROADMAP PLANNING	29
6 CONCLUSION	31
6.1 NEXT STEPS.....	31
7 ANNEX 1: SCRUM IN A NUTSHELL.....	33
7.1 ROLES AND ACTORS.....	33
7.2 AGILE TOOLKIT	35
8 ANNEX 2: EGI.EU COLLABORATION TOOLS OVERVIEW	43

TABLE OF FIGURES

Figure 1: Principal inter-work package information flow.....	6
Figure 2: Inter-work package collaboration between WP3 and WP5	27
Figure 3: The key roles in an agile project.....	34
Figure 4: Consecutive sprints continuously expand and improve a product, steered by the Product owner	38
Figure 5: Dynamics and KPIs in an agile project	41

TABLE OF TABLES

Table 1: Stakeholders of WP5 activities	10
Table 2: Chronological planning of WP5 activities	12
Table 3: Initial toolkit selection in DCH-RP WP5	15

1 EXECUTIVE SUMMARY

The partners in Work Package 5 will conduct a number of Proofs of Concept (PoC) to answer a number of questions that arise from the work on the roadmap for the preservation of digital cultural heritage – the heart of the output of the DCH-RP project, developed and published by Work Package 3.

Those questions are not known a priori, even though a set of topics for examination in the PoCs is provided in the introductory section. However, these questions will broadly fall into the following categories:

- Functionality of the preservation infrastructure components;
- Non-functional capabilities (e.g., scalability, reliability) of the examined e-Infrastructure;
- Acceptance and usability of services;
- Sustainability and technology insertion options into local and international DCH infrastructures.

To prevent over-engineering and over-planning in the beginning, the scope and actual content of this document are limited by design. It rather describes the methodology and tools with which the planning and execution of the Proofs of Concept will be carried out. Following the Scrum agile project management methodology, this deliverable briefly looks at the main elements of Scrum, their intent and how they will be used in this Work Package, and the way they will manifest and be maintained in existing collaborative tools.

Through providing description of each PoC partner's background, context and objectives, this document establishes a common mindset particularly for the partners through cross-review and discussions. Even though a common interest and good overlap in common objectives is assumed, we nonetheless expect differences in objectives, skillsets and data that needs to be preserved. By explicitly providing each partner's position, we establish an open collaboration and facilitate (although as a first step) forming a team of partner representatives as the nucleus of the project team at the heart of the agile Scrum process.

With identifying commonalities in skills, background and objectives we will have a starting set of topics to work with right away.

By accepting that we do not know everything beforehand, and appropriately adapt our way of how we are going to conduct the planning and execution of Proofs of Concept, Work Package 5 allows for very flexible planning and collaboration with all other partners and interaction points in the project.

Finally it should be noted that WP5 was introduced in the project plan only during the negotiation to reinforce and give evidence to the results of the project by experimenting and putting into practice the concepts that will be established in the roadmap. Moreover the activities that are planned in WP5 are strictly related to the results of WP3 and WP4. This is why this document is to be considered the first step of an ongoing process and will evolve and be kept updated throughout the duration of the Proof of Concepts, taking into consideration the inputs coming from WP3 and WP4.

2 INTRODUCTION

As stated in the Description of Work for DCH-RP, preservation is one of the most challenging problems of the current digital era, applying to all sectors of society, including the DCH sector. Preservation is a broad concept, but DCH-RP defines preservation as the combination of preserving:

- Data (digitized and born-digital content like databases, catalogues, files, etc.) and
- Information associated with that content (so-called ‘infostructure’, referred to also as metadata).

DCH-RP deals both with ‘long-term preservation’ (preserving for an unpredictable long period of access and use) and ‘short-term preservation’ (preserving for a relatively short period of access and use). The main objective to be achieved by the project is to design a sound roadmap for the implementation of an e-Infrastructure for preservation of DCH content, as part of a more general vision towards and Open Science Infrastructure for DCH in 2020.

The project will organize a number of Proofs of Concept (PoCs) where cultural institutions and e-Infrastructure providers will experiment with the actual use of grid and cloud services to store cultural digital resources. These PoCs will run in the following countries:

- Italy, with e-Infrastructure facilities provided by INFN and data provided by ICCU;
- Poland, with e-Infrastructure facilities and data provided by PSNC;
- Hungary, with e-Infrastructure facilities provided by NIIFI and data provided by Hungarian cultural institutions that will cooperate with NIIFI;
- Sweden, Belgium and Estonia, with e-Infrastructure facilities provided by the NGIs exploiting the Memorandum of Understanding signed in the frame of DC-NET and data provision coordinated by the DCH-RP partners.

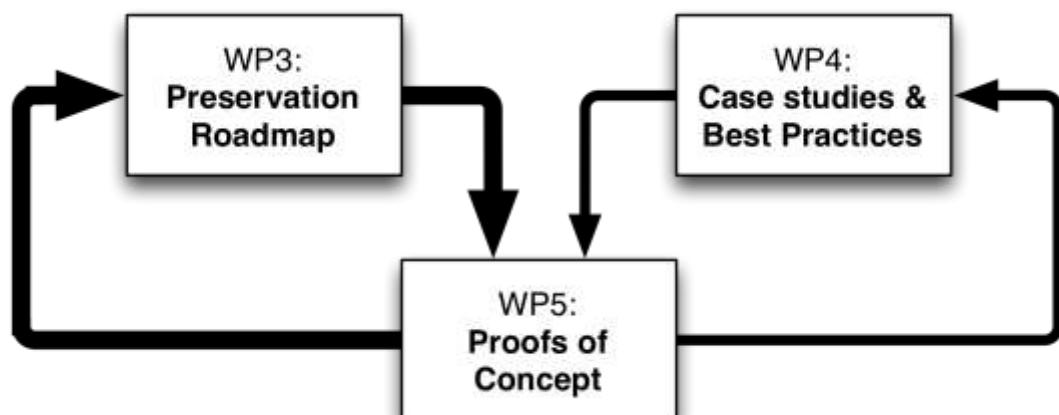


Figure 1: Principal inter-work package information flow

Work Package 5 has been designed to coordinate and carry out these PoCs, and the present deliverable provides the planning of WP5 activities to be executed by the

involved partners. The organization of the PoCs is outlined in this document and based on information given by each of them regarding, for example:

- Selected content to be used for the test;
- Description of the content that will be used for the testing (text, bibliographic records, image repositories, audio-visual repositories, etc.);
- Institutions involved and contact persons;
- If the content used for the testing is open or protected;
- Approximate quantity of content that will be used for the testing.

The first PoC will take place in months 6 – 12, after the delivery of D3.1 from work package 3. The results will be documented and analyzed in a detailed report (D5.3).

The second PoC will be carried out in month 15 – 21, after the delivery of D3.2, and the results documented and analyzed in D5.4.

2.1 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 planning for the PoCs will establish the methodology by which WP5 will conduct its work. It will also provide the direction of each PoC in terms of strategic outcome and contribution to the preservation roadmap: While the first PoC is dedicated to provide a gap analysis of *existing* national preservation infrastructure, the second PoC will attempt at implementing an international distributed preservation e-Infrastructure across WP5 partners. 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.

2.2 STRUCTURE OF THE DOCUMENT

The remainder of this document is divided into two further sections and two annexes, as follows.

Section 3 briefly discusses alternative activity management models and the WP5 partners' decision to employ the Scrum¹ methodology, and how the respective tools and processes will be implemented with the existing collaboration infrastructure.

Section 4 describes the partners' expertise and role within their local partner network and DCH community, an outlook on the context and constraints with each partner. Lastly, each partner expresses their individual objectives that they aspire to reach through their contributions to the PoCs.

Based on sections 3 and 4, section 5 outlines the technical planning for the two Proofs of Concepts that will be conducted in WP5 over the course of the project.

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

¹ <http://www.scrum.org/ScrumGuide.aspx>

Lastly, Annex 1 (section 7) provides an overview of the Scrum methodology that will be employed in this Work Package, and Annex 2 (section 8) describes the collaborative tools available for the DCH-RP project.

3 METHODOLOGY, TOOLS AND PROCESSES

3.1 AUDIENCE AND STAKEHOLDERS

The DCH-RP project aims to play an important role in taking forwards the adoption of e-Infrastructure as a computing and storage platform by the DCH research community. The following table describes the stakeholders for the 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/P ²	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	O	Prospective users of the new services	Usefulness and usability of the new services, range of	Medium

² CH institutes are predominantly observers even though a selected few are participating in the PoCs in this project.

Stakeholder	Participant/ Observer	Role	Area of interest	Importance of information
community			services	
e- Infrastructure providers	O/P ³	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: Stakeholders of WP5 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 referenced 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. Since Work Package 2 is devoted 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 among people*; collaboration tools and methodology should satisfy this prime requirement.

3.2 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.

³ The same applies to e-Infrastructure providers. While some are participating in the PoCs, most of them are predominantly observers.

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.

3.2.1 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 deliverables. It is for this reason that most of the milestones are timed in advance of the corresponding 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 1 st report												Yellow			
1 st 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 2 nd report												
2 nd PoC report										D5.4		

Table 2: 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

3.3 AGILE PROJECT MANAGEMENT VS WATERFALL MODEL

Classic project management organizes work in well-defined sequential phases. Commonly known as the “waterfall model”⁴ 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 results.

In the real world, however, and particularly in science projects of exploratory nature such as DCH-RP, not all preconditions and requirements of the overarching objectives are known beforehand. 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 wherever required, and to adjust the direction of the project or some of its activities.

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. The partners involved in Work Package 5 decided to implement the Scrum agile management methodology over the course of the project (see also Annex 1 in section 7).

⁴ http://en.wikipedia.org/wiki/Waterfall_model

3.4 IMPLEMENTING SCRUM IN WORK PACKAGE 5

The Scrum⁵ agile methodology is extensively described in numerous books and online media⁶. Annex 1 (see section 7) provides an overview of Scrum for the sake of completeness of this document; those unfamiliar with Scrum may want to read Annex 1 before reading further.

The project partners also anticipate the need of components integrated into the preservation infrastructure that enable collaboration of contributing institutes on the task of preservation, but also end users that include digital cultural heritage into their own research. EGI.eu, as a contributing partner of Work Package 5, already provides an integrated set of collaboration tools that it made available for the DCH-RP project free of charge⁷. Thus, the DCH-RP project will use these tools on two levels:

- As a DCH-RP project collaboration infrastructure;
- As a set of collaboration tools considered for the preservation roadmap.

Annex 2 (see section 8) provides an overview of the EGI.eu collaboration tools. Work Package 5 decided to use some or all of the available tools to implement the Scrum methodology as described in the remainder of this section.

3.4.1 Assigning Scrum roles and actors

In Scrum, four roles work together to deliver, at the end of the project, the initially envisioned product (see Annex 1, section 7). Considering the structure of the DCH-RP project, and WP5 in particular, the partners will set out with the following assignments:

- **Facilitator**

Being a fairly obvious assignment, EGI.eu, as the Work Package 5 leader, will assume the role and responsibilities of the Facilitator.

The named individual is Michel Drescher (EGI.eu).

- **Product Owner**

The main contributor to WP5's work will be Work Package 3, augmented by material from Work Package 4. Given that project partners agreed to introduce the agile methodology, only one named individual should act as the Product Owner. Hence the leader of Work Package 3, Riksarkivet (RA), shall be the Product Owner.

The named individual is Borje Justrell (RA).

- **Project Sponsor**

Ideally, the Project Sponsor would aggregate all necessary empowerment. However, in this project, reality provides us with one financial sponsor (the European Commission), the team member sponsors (the partner institutes participating in Work Package 5, and the management sponsor (the Project Coordinator together with the Project Management Board). However, a Technical Coordinator for the project was proposed and accepted for the DCH-RP project,

⁵ <http://www.scrum.org/ScrumGuide.aspx>

⁶ See for example this Google search: <http://goo.gl/bCW7v>

⁷ The tools are provisioned free of charge for the duration of the DCH-RP project given that no modification of the current deployment are required.

providing a fairly natural choice for this role. Thus, the Technical Coordinator of the DCH-RP Project, Promoter SA (PM), shall represent the Project Sponsor for WP5.

The named individual shall be Antonella Fresa (PM).

- **Project Team**

Last but not least, the Project Team will comprise all the partners collaborating in the Proofs of Concept. Again, the situation needs to take into account that the Project Team members will not be available 100% of their time, as the Scrum guidelines recommend. Rather, the partners act as representatives to their local institutions and colleagues who will conduct the actual work.

Thus, the members of the Project Team will be named individuals representing the work package partner institutes as follows:

<u>ICCU</u> :	Giovanni Ciccaglioni, Patrizia Martini, Laura Ciancio
<u>INFN</u> :	Roberto Barbera
<u>RA</u> :	Sanja Halling, Rolf Källman, Eva Toller
<u>BELSPO</u> :	Rosette Vandenbroucke, Hilke Arijs, Hans Opstaele
<u>EVKM</u> :	Indrek Eensaar
<u>PSNC</u> :	Maciej Brzezniak, Norbert Meyer
<u>NIIIFI</u> :	Lajos Balint

3.4.2 The Scrum toolkit in Work Package 5

Re-using the existing collaborative tools provided by EGI.eu, most of the contemporary Scrum tools and processes could be accommodated. [Table 3](#) provides an initial mapping of collaboration tools to Scrum artifacts. This list will be continuously reviewed and adjusted over time wherever and whenever required.

This initial selection does not imply that the collaboration is restricted to the use of the indicated tools. In fact, [Table 3](#) provides only a subset of the tools EGI.eu has made available (with the A/V facility being provided by INFN). Annex 2 (see section 8) provides an overview of the tools EGI.eu provides for use in the DCH-RP project and possibly beyond.

User stories and epics are expected to evolve in a manageable frequency before they are handed over to Work Package 5. The same applies to tasks even though these are likely to keep being maintained and changed until they are completed. However, each must carry unique IDs to be able to correlate them to information used in the backlog, which will be initially kept and maintained in an Excel spreadsheet. To easily correlate tasks, epics and user stories, their IDs and titles will be used in the backlog, but no other information besides task metadata.

The backlog is a pivotal tool in Scrum project. It requires a fair amount of calculation and automation to allow the facilitator to focus on the issues at hand. Out of the available tools, Excel spreadsheets seem the best choice at this point in time. As much as necessary, and as little as possible, of information duplication between the information in the Wiki and the Excel document will have to be achieved. The exact balance is difficult to determine beforehand, but should follow the general guideline of using the spreadsheet as a working tools, and the Wiki as the more publicity-oriented and more collaborative tool.

Scrum artifact	Wiki	Offline documents	A/V conf.	Doc. DB	Mailing list	Meeting planner
User Story	X					
Epic	X					
Task	X					
Backlog		X		X		
Sprint planning			X			X
Sprint	X	X		X	X	X
Sprint review	X	X	X			X
Sprint retrospective			X			X
Reporting	X	X				

Table 3: Initial toolkit selection in DCH-RP WP5

During the sprint, the project team may use any of the provided tools as they see fit. In fact, many contributions from WP3 and WP4 will come through any of the collaboration channels indicated above. No limit is imposed on the team under the condition that all information is publicly accessible and permanently documented (unless specific conditions require confidentiality).

Sprint planning, review and retrospective meetings are planned to use audio/video infrastructure (currently, Adobe Connect Pro provided by INFN) with screen sharing functionality to be able to provide a close to interactive view on the backlog, Wiki and any other information during the meetings.

In conclusion, Work Package 5 will begin to work with these tools, and use the sprint retrospection and review meetings to continuously monitor whether the toolkit needs to be adjusted to the needs of the team. While principally all tools are available for any intent in WP5 (and any other work packages in the DCH-RP project), [Table 3](#) indicates the expected predominant use of the tool for the indicated activity in WP5.

4 STAKEHOLDER BACKGROUND AND OBJECTIVES

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 they have. 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 project partners are likely, although sufficient overlap in common goals is expected. However, it is important to repeat that all partners in the Proofs of Concept are targeting three types of user communities, as initially described in the project's DoW:

- 1) Content providers;
- 2) Policy makers and program owners;
- 3) End users accessing the resulting DCH infrastructure to access data that content providers make available for subsequent research.

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.

4.1 ITALY (INFN, ICCU)

4.1.1 Pilot lead partner

The Central Institute for the Union Catalogue for the Italian Libraries and for Bibliographic Information (ICCU) promotes and coordinates cataloguing and 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.

The INFN (Istituto Nazionale di Fisica Nucleare) is the Italian National Institute of Nuclear and Particle Physics (www.infn.it). Founded in 1952, INFN is a governmental research organisation, which promotes, co-ordinates and funds nuclear, particle and high-energy physics research. INFN staff research personnel amounts to more than 1900 peoples with a equivalent number of associates from Universities and other Scientific National Institutes.

INFN has a long tradition of dissemination and training, dedicated both to the scientific community and to the general public. INFN runs a country-wide Grid infrastructure for e-

science and, since the year 2000, has taken part as a stakeholder in many European EU co-funded Grid Projects.

In the context of DCH-RP, and WP5 in particular, INFN will provide storage resources and the e-Culture Science Gateway.

4.1.2 Context and constraints of the pilot

Cultural institutions can become partners of Internet Culturale and have their own digital archive. In this data federation model, each institute maintains their own digital archive (comprising of manuscripts, books, digital facsimiles of music scores, serials and periodicals) augmented with local metadata that is produced according to national and international standards (MAG, METS, PREMS). All local metadata are indexed by Internet Culturale's services through the standard protocol OAI-PMH⁸. 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;
- Speed;
- Migration to new formats;
- Documentation.

4.1.3 Objectives

For our digital archive, the MagTeca, 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);

⁸ Open Archive Initiative (OAI) Protocol for Metadata Harvesting (PMH) – see <http://www.openarchives.org/pmh/>

- 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).

4.2 SWEDEN (RA)

4.2.1 Pilot lead partner

Digisam is a secretariat for National coordination of digitization, digital preservation and digital access to cultural heritage. In order to coordinate the continued development work on digitization 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 digitization, 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 digitization.

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.

4.2.2 Context and constraints of the pilot

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.

4.2.3 Objectives

The first PoC in Sweden will focus on archival data and testing will be expanded to data from other cultural heritage institutions in the second PoC. The objective of PoCs is to test preservation of cultural heritage data thought services provided by e-infrastructures in order to explore requirements and new solutions for sustainable common and cost effective long-term preservation of cultural heritage data.

Still, for our PoCs only the first group of users "cultural heritage institutions will be concerned. Even if there is a description of three user communities in DoW, those communities can be more relevant in different parts of project. As we see it, the most relevant user group for WP5 is the first group of users.

4.3 BELGIUM (BELSPO)

4.3.1 Pilot lead partner

Belspo, the Belgian Science Policy, is the DCH-RP project partner responsible in WP5 for the organization of the Proofs of Concept. Belspo is not a cultural heritage institution 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 institutions that showed their interest in participating actively in the project are:

- The Royal Institute for Arts (KIK);
- The Royal Museum for Arts and History (KMKG);
- The Royal Library (KB);
- The State Archives (RA).

The data of KIK and KMKG are of the same type (images or artwork in 2D and 3D, multimedia) while KB and the RA have digitized documents.

4.3.2 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. Furthermore, the eCSG will be deployed and tested in the Belgian environment.

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.

4.3.3 Objectives

Key issues for the Belgian participants are 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.

4.4 ESTONIA (EVKM)

4.4.1 Pilot lead partner

The Estonian Ministry of Culture (EVKM) is the project partner responsible for the organization of Proofs of Concept in WP5. EVKM is not a cultural heritage organization itself and neither does it manage collections. The 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. Kanut serves most Estonian museums and digitizes a large variety of objects. It is the first cultural organization in Estonia to engage in 3D digitization of cultural heritage. Due to this, Kanut has a broad collection of different types of digitized objects and data.

4.4.2 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. The 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. In Estonia, EENet provides various services to educational, cultural and science organizations. So far, EENet has not provided the service of long-term preservation of cultural heritage. The Conservation Center Kanut lacks the experience of preserving data using cloud-solutions. EENet can provide up to 100 TB of its cloud-solutions data storage space for pilot-solutions. In order to ensure the success of the project, it is imperative to make sure that the e-CSG corresponds to EENet infrastructure.

4.4.3 Objectives

The key issue for the Conservation Centre KANUT is to test different types and format of 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. The Conservation Centre KANUT is also interested in long-term cooperation with the EENet

in order to solve the issue of long-term data preservation. They are also interested in cooperation with the EENet regarding 3D modeling. Cooperation concerning 3D models for data storage in EENet would be an excellent starting point for continuous cooperation. The objective of EENet is to gain experiences in using e-CSG and in preservation of cultural heritage.

4.5 POLAND (PSNC)

4.5.1 Pilot lead partner

The 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 PIONIER⁹ 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 Storage¹⁰, PLATON¹¹ as well as EuropeanaCloud¹², LoCloud and SYNAT¹³, IMPACT¹⁴, SUCCEED and Wf4Ever¹⁵.

Digital Library of Wielopolska¹⁶ (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 Libraries¹⁷ (PFSL). PSNC has very good contacts with DLoW and PFSL members and will exploit them for the DCH-RP PoC purposes.

4.5.2 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

⁹ Polish Optical Internet. <http://www.pionier.net.pl/online/en/>

¹⁰ National Data Storage. <http://nds.psnc.pl>

¹¹ PLATON - Science Services Platform. <http://www.platon.pionier.net.pl/online/?lang=en>

¹² <http://europeana.eu>

¹³ <http://www.synat.pl>

¹⁴ IMPACT Competence Centre. <http://digitization.eu>

¹⁵ <http://wf4ever-project.org/web/guest/home>

¹⁶ <http://www.wbc.poznan.pl/dlibra?action=ChangeLanguageAction&language=en>

¹⁷ <http://www.pfsl.poznan.pl/en>

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 many VOs. For instance, the DCH-RP VO vo.dch-rp.eu 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 UDPR, 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.

4.5.3 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/dArcheo/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.

4.6 HUNGARY (NIIIFI)

4.6.1 Pilot lead partner

NIIIFI (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 NIIIFI in DCH-RP, as major representative in the country of performing DCH collection and preservation.

NIIIFI is the NREN (National Research and Education Networking organization) in Hungary and also serves as NGI of the Hungarian research and education community. NIIIFI 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. NIIIFI 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.

4.6.2 Context and constraints of the pilot

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

- N x 10 Gbps hybrid network (IP + e2e);
- 3200 km DF (CEF) backbone (\rightarrow 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 authorization) 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 NIIIFI cloud features for various application areas (including archives) are listed here:

- NIIF IaaS Cloud = distributed resource management + virtualized machines + network;
- Private cloud for NREN users;
- OpenNebula + KVM - open source;
- Web management 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 NIIIFI (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.

4.6.3 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.

5 PROOFS OF CONCEPT PLANNING OVERVIEW

With the planning methodology put in place in section 3, and an initial overview of the local assets provided in section 4, this section will provide the large-scale planning and objectives for the two Proofs of Concept planned in DCH-RP.

In essence, WP3 and WP5 will collaborate in a virtuous cycle, with intermediate outcomes of either work package informing and influencing the subsequent next phase of the other work package (see [Figure 2](#)).

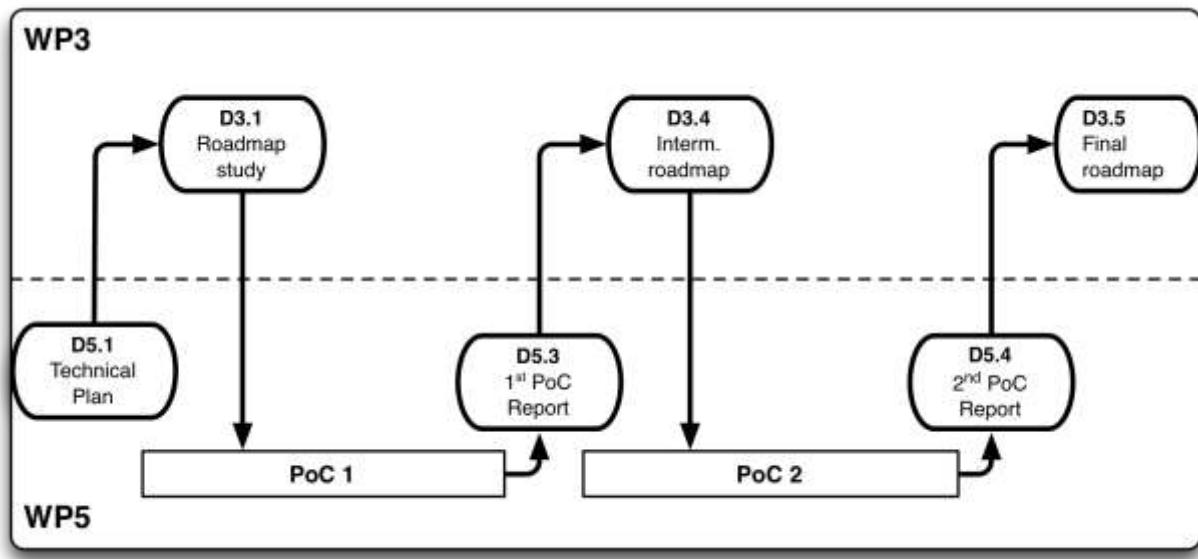


Figure 2: Inter-work package collaboration between WP3 and WP5

Although not shown in [Figure 2](#) for brevity and clarity, D5.1 and D5.3 will form the basis of the subsequent Proofs of Concept. The concrete and detailed planning for these PoCs will be worked on and documented through the sprint planning and task descriptions provided in the DCH-RP project wiki.

5.1 PROOF OF CONCEPT EXECUTION PLANS

As described in section 3.2.1, WP5 will conduct two Proofs of Concept, and document the outcomes for further use in the project as briefly described above. While both Proofs of Concept are considered to last for 8 months each, the first Proof of Concept bears the risk of the upgraded eCSG (task 5.2) to be still ongoing (D5.2 is expected to be ready at month 6).

The major challenge in Work Package 5 is that the activities that are planned in this Work Package are strictly related to the results of WP3 and WP4, whose first results will be available in Month 6 (D3.1). This brings the project into the situation that activities for a Proof of Concept have to be planned and implemented (as manifested in this deliverable D5.1) *before* actually knowing what needs to be proven and against which criteria the proof runs have to be executed against. This is why this document is to be considered the first step of an ongoing process and will evolve and be kept updated

throughout the duration of the Proof of Concepts, taking into consideration the inputs coming from WP3 and WP4.

Considering this context, the partners in not only agreed to employ the Scrum methodology for planning purposes (see section 3.3), but also to organize the two Proofs of Concept being themed as follows:

Proof of Concept 1:

“Gap analysis of existing infrastructure against an evolving roadmap”:

- a. Inventory of existing infrastructure;
- b. Identify preconditions to a coordinated gap analysis;
 - i. Technical infrastructure;
 - ii. MoUs between the project and participating NRENs;
 - iii. Digital data available for experimentation;
 - iv. Technical methodology;
- c. Conduct gap analysis:
 - i. Continuous synchronization with WP3 for adjustment as required;
- d. Final report of PoC 1 (D5.3).

Proof of Concept 2

“Validate the preservation roadmap against an integrated preservation e-Infrastructure in Europe”:

- e. Bring together D5.3 (gap analysis), current roadmap (D3.4) and ancillary input from WP4 (Tasks 4.1, 4.3, and 4.4);
- f. Run the second gap analysis:
 - i. Continuous synchronization with WP3 for adjustment as required;
- g. Final report of PoC 2 (D5.4).

Of course, most of the information necessary for Proof of Concept 2 will be only available over the course of the project. Hence any information and visions and goals described for Proof of Concept 2 will have to be taken with great care due to the limited availability of information.

For PoC 1, however, the following activities will be further elaborated in the short term, building on available team skill sets:

Sweden (RA) with a long history and experience in preservation may bring in the archival data, metadata description, formats and standards, and share these with the other partners in Work Package 5. Given the success as planned, it may then expand on federating in data from further cultural institutes in Sweden and elsewhere in Europe.

Belgium (BELSPO) may focus on providing cultural data using the Belgian Grid infrastructure provided by BELNET, and verify that approach against the evolving preservation roadmap.

Estonia (EVKM) may use KANUT and their experience in data archiving processes, in a scenario to transition to EENet as service provider. This will bring valuable experience and lessons learned in a transition exercise that all partners may face during the course of developing and implementing the preservation roadmap within and beyond the DCH-RP project.

Poland (PSNC) already provided the most developed PoC scenario applicable for Poland in section 4.5. Together with **Italy (INFN, ICCU)**, both countries may explore international aspects of a preservation e-Infrastructure and data federation.

Hungary (NIIFI) may prove how the roadmap validates against the requirements from the aspect of (international) networking infrastructure provisioning.

5.2 FEEDBACK TO PRESERVATION ROADMAP PLANNING

The peculiar project situation mentioned above has one benefit. Very much in line with the Scrum methodology, the preparation of D5.1 brought the project partners together and engaging in fruitful discussions stretching our beyond the scope of Work Package 5.

During those discussions the following topics were brought up for potential inclusion in the preservation roadmap and subject to verification during the Proofs of Concept. They are provided in the current form as initial feedback from Work Package 5 to Work Package 3 for consideration.

These topics need to be broken down into smaller, actionable items for inclusion into the roadmap. We expect that these topics will “return” to Work Package 5 in the form of a continuous stream of User Stories and other background material.

This activity taking place in WP3 usually requires some forerun before the continuous supply of material is established. In the meanwhile the Project Team in WP5 can already work on the topics mentioned above in section 5.1., and start preparing the generic parts of the Proof of Concepts without risking too much diversion from the intended direction.

5.2.1 Functional requirements:

1. Federation:
 - a. AAI (such as re-use local institution's authentication mechanisms);
 - b. Federated data archives and metadata services;
 - c. Central access gateway/portal vs. federated access;
2. Integration:
 - a. Grids, Clouds, and local preservation services;
3. Support:
 - a. Models, infrastructure, responsibilities;
4. Data retention/preservation:
 - a. Raw data formats and standards;
 - b. Metadata retention & standards;
 - c. Technology insertion strategies to replace obsolete with new technology;
5. Data curation:
 - a. Preservation QA;
 - b. Digital archive maintenance QA;
 - c. Obsolete (“dark”) data;
 - d. Data provenance;
 - e. Data ownership;
6. Data quality:
 - a. Bit-level preservation;

- b. Preservation outcome quality;
- c. Integrity checks.

5.2.2 Non-functional requirements:

1. Taxonomy:
 - a. Data classification;
 - b. Data quality (e.g. eligibility for preservation, cost/benefit analysis);
 - c. Glossary;
2. Interoperability:
 - a. With existing preservation archives and tools;
 - b. Between tools on the infrastructure layer;
 - c. Existing standards, necessary standards, and profiling options;
3. Sustainability:
 - a. Financial (TCO, CAPEX vs. OPEX, etc.);
 - b. Technical (tool choice, tool provider choice, etc.);
4. Technical architecture:
 - a. Common/shared tools vs. local tools and tooling;
 - b. Responsibility for maintenance, and technological evolution;
 - c. Critical vs. non-critical components;
5. Community organization:
 - a. VRE, VOs and coordination.

6 CONCLUSION

As explained in the introduction, WP5 was introduced in the DoW at the time of the negotiation to reinforce the outcome of the project. The planning of this new WP5 was conceived as a support for the validation of WP3 and WP4. In addition, in order to overcome the physiological inertia of the initial phase of each project, the delivery of D5.1 has been put very early. Thus, for Work Package 5, the first three months of the project represented a challenge to accomplish what was planned (also in consideration of the impediment due to the change of partner representative at EGI.eu). The EC agreed to accept the delivery of D5.1 by January 2013. In the course of this initial period, partners in WP5 swiftly started discussing among each other not only about issues within WP5, but across all work packages, as originally foreseen in the planning of the interaction and iterations between WP3, WP4 and WP5. With that constructive atmosphere, WP5 was able to contract all necessary contributions and to collate them into this deliverable D5.1.

Through these discussions, it became very clear that the interactions between WP3, WP4 and WP5 are indeed iterative in their nature, both on the macro-led deliverable-oriented project structure and on the day-to-day interaction between project partners.

Section 3 illustrates the methodology, tools and artifacts that will be employed in Work Package 5, together with the rationale behind this decision. While this covers in sufficient detail *how* we will accomplish our work, it leaves behind *what* we are going to do – it does not describe the *product* or, more specifically, what the product of WP5 consists of. And without knowing our product, we will not have any direction to go into, and we will fail in demonstrating progress.

Quoting the project DoW, “WP5 aims to proof [sic] in concrete experiments the concepts established by WP3.” This is our goal and our product. Section 4 describes initial views of the Proofs of Concept as provided by the work package partners, already illustrating how the tasks in the sprints may be conducted, and who the target audience of this product will be (i.e., all three identified user communities).

Clearly, this is not a software development project, thus our product naturally will not be a piece of software. Although the supplied User Stories are expected to describe functionality of software, the resulting task will not be to implement that functionality. Rather, it will have to deploy software and assess whether it satisfies the requested functionality and requirements, or not. The results of these tasks may be negative or positive or impossible to assess. It is thus important to realize that the Product Owner’s decision whether to reject or accept a task in the sprint review must not be based on the existence of a functionality, but whether the task outcome properly assesses the requested functionality by the common criteria defined in section 5.

6.1 NEXT STEPS

The next immediate actions need to cover a number of topics to get the deployment and testing of the Proofs of Concepts going.

- **Administrative**

The initial role assignments must be confirmed and briefed in their roles and responsibilities. Inaccurate initial assignments need to be corrected and new members introduced to the team and the agile methodology.

- **Agile tool chain**

Facilitator and Project Team need to quickly agree on the collaborative tools they wish to use for their work. For all Project Team-internal work, any tool is acceptable that all team members commit to; tools and documents that interface between Facilitator and the Team, the Product Owner and the Project Team need to be discussed brief, but in detail.

Once an agreed tool chain is available, all participants need to start using them. An initial list of tools and their intended use is suggested in section 3.

- **Project input and material**

To start working effectively, Work Package 5 needs a continuous stream of input coming from Work Package 3 and, to some lesser extent, from Work Package 4. These communication channels need to be set up as soon as possible, so that Work Package 3 can start supplying User Stories to Work Package 5.

In parallel, the Project Team can already start collecting tasks for activities that are fairly obvious even without input from Work Package 3.

7 ANNEX 1: SCRUM IN A NUTSHELL

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”, regular “stand up meetings”, “user stories” 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.

The following subsections provide an overview of the Scrum methodology. It is intended to provide ancillary information within this deliverable for those that are unfamiliar with the Scrum methodology.

7.1 ROLES AND ACTORS

In agile-managed projects, more specifically projects employing the Scrum methodology, 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. [Figure 3](#) provides an overview and the key interaction network between the roles in an agile managed project. It is important to note that the terms are kept generic¹⁸ illustrating that the Scrum methodology is applicable to many domains where agile project management is feasible to implement. The following subsections explain in more detail the different roles in Scrum project management.

7.1.1 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 prioritizes 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

¹⁸ Often, Scrum projects name the roles how it is written in many Scrum guidelines. Coming from the Software engineering domain, the originally coined terms tend to (a) mislead the casual observer and (b) limit the agile methodology to software engineering only.

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.

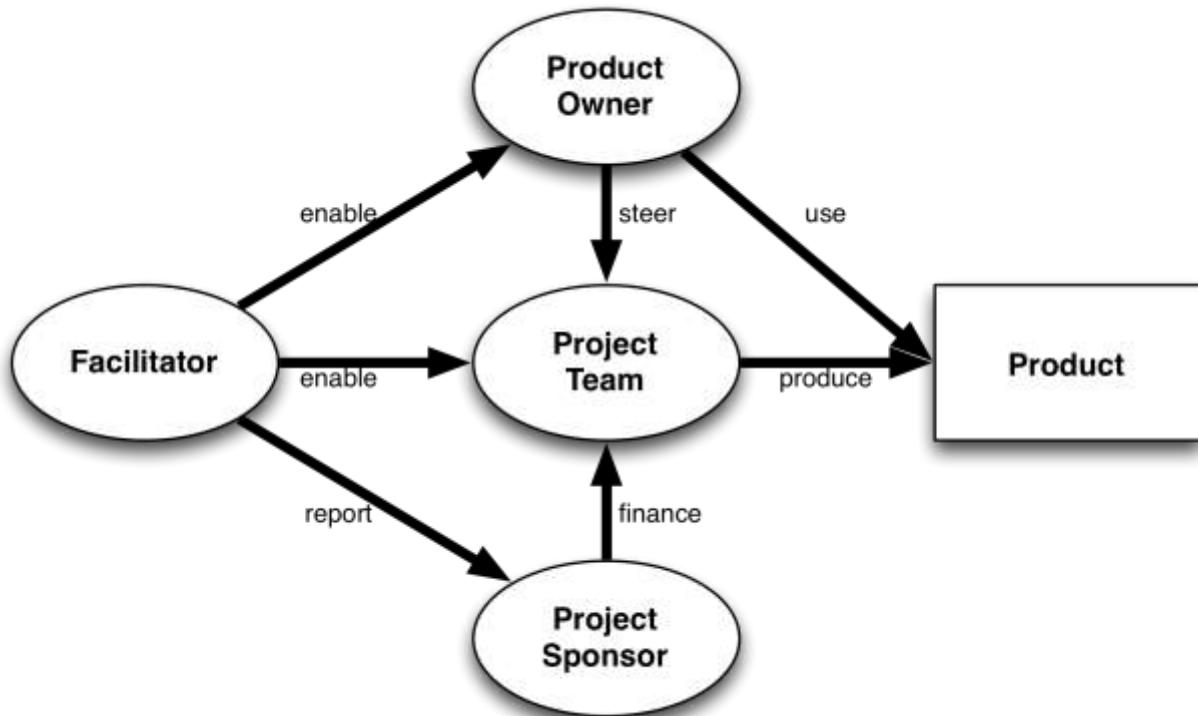


Figure 3: The key roles in an agile project

7.1.2 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.

7.1.3 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.

7.1.4 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.

7.2 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. The stakeholders, and more importantly the project team, need to decide whether to adopt the Scrum methodology and how to implement it.

7.2.1 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 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 *publicize* 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.

7.2.2 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 prioritization by the consuming stakeholders influencing and steering how the Project Team will conduct its work. Epics, if employed, are often a collaborative work of the product owner and the team.

7.2.3 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 most 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.

7.2.4 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;
 - Prioritization;
- 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) ;
 - Prioritization 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.

7.2.5 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).

Sprints have a fixed structure in agile project management (see [Figure 4](#)) 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).

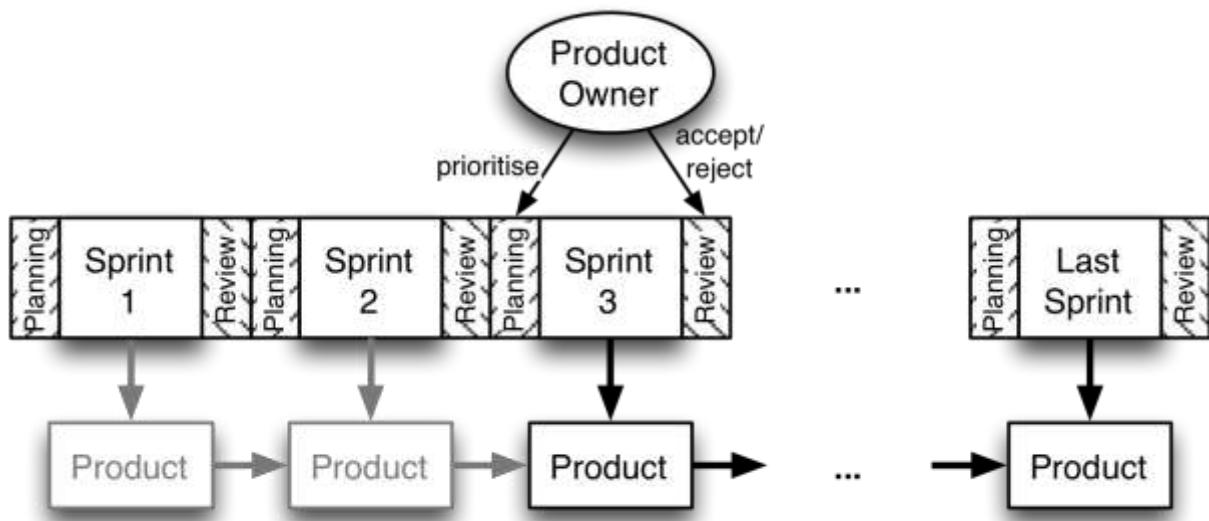


Figure 4: Consecutive sprints continuously expand and improve a product, steered by the Product owner

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:

a) 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.

b) 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-organizing 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,

c) 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, Prioritization, 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.

d) 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. Frequently, the retrospection is scheduled back-to-back with the review, or folded into the review with everyone else leaving the room except the required participants.

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 recorded¹⁹, 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.

7.2.6 Management, reporting and forecast

The dynamics of an agile managed project need to be captured, and the resulting figures (mainly) used for project management, risk management and projection. [Figure 5](#) illustrates how the dynamics and processes in an agile project influence the progress of the project. The progress is captured using five key performance indicators (KPI) that are described in more detail below.

¹⁹ The Facilitator may provide templates and storage location for the material, but it is the Product Team's responsibility to record the discussed topics.

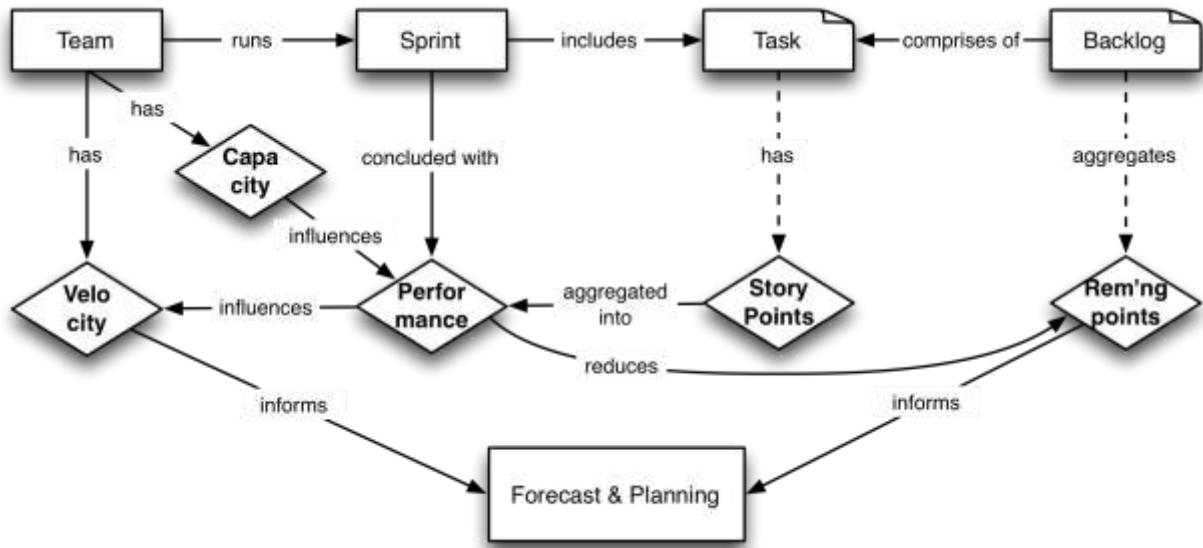


Figure 5: Dynamics and KPIs in an agile project

a) Team capacity

The team capacity is usually assumed to be 100%, i.e. all team members are available at all times during a sprint. In reality, however, this is not always true, hence the team capacity needs to be determined for a sprint in the sprint-planning meeting (or earlier). The capacity can be less, or more than 100%. For example, if a team member may get ill, goes on (well deserved) vacation, or is otherwise not available for the team during the sprint, the capacity of the team will be lower for that sprint and any subsequent sprint until that person returns into the team. On the other hand, if the project sponsor decides to temporarily increase the team size, the facilitator needs to reflect this in an excessive capacity figure so that the team, after returning to its original size, will not be overcommitting tasks into the next sprint due to an artificially inflated team velocity.

b) Story points

The Project team discusses each received user story (thus turning it into a task), and assesses each task's necessary effort with a number of story points. This abstract number often reflects the expected work hours, even though it should not. A sprint is always planned so that the aggregated story points, or the expected performance, will not exceed the current team velocity (see below).

c) Sprint performance

In the sprint review, the Product Owner decides whether to decline or accept the Project Team's delivery. That is, the Product Owner simply decides, according to the task's "definition of done" (see above), whether the task is done or not. This is repeated for every task that the Team has decided to deliver (i.e. mark as "delivered"). Once all Tasks are assessed, all rejected tasks are added to the subsequent sprint. All accepted tasks are marked as accepted, and all associated story points aggregated into a sum – the **sprint performance**. Advanced agile projects take the team capacity into account, and normalize the sprint performance so that it indicates the performance as *if the team had been at 100% capacity*. Even

through this normalized value is identically named, it is relatively easy to determine which performance is tracked by searching the backlog and related material for any mention or figure of team capacity. The *normalized* sprint performance is calculated through dividing the aggregated story points by the team capacity for that sprint.

d) Velocity

The team velocity is calculated as the arithmetic average over the sprint performance values of the last three consecutive sprints. The exact number of sprints taken into account depends on the agreement in the Project Team, the Facilitator and the Project Sponsor.

The Facilitator keeps tract of the team's current *and* historic velocity, as these figures are required for forecast and trend reporting towards the Project Sponsor.

e) Remaining (Story) Points

Over the course of the project, the backlog will grow with many tasks, thus increasing the total number of story points captured in the backlog. This number will grow and shrink according to the number of tasks in the backlog, accepted task deliveries and re-estimating task effort. At the end of a sprint, the Facilitator calculates the number of remaining story points simply by removing the accepted tasks from the equation of all tasks in the backlog.

The remaining story points thus illustrate the remaining effort until the aspired product is completed.

Based on these KPIs the Facilitator, the Product Owner and the Project Sponsor have an overview at all times whether corrective actions and further prioritization are necessary or not.

8 ANNEX 2: EGI.EU COLLABORATION TOOLS OVERVIEW

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 project²⁰. 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:

- a) Provide the project work packages and members with a powerful set of collaboration tools, and
- b) 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.

8.1.1 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> for anyone with an SSO account. SSO group owners manage SSO groups and membership through the same address.

8.1.2 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

²⁰ 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.

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.

8.1.3 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.

The DCH-RP specific Wiki is accessible at the address https://wiki.egi.eu/wiki/DCH-RP:Main_Page²¹.

8.1.4 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 document²². 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 project²³.

8.1.5 Meeting planner

A meeting planner allows for registering and scheduling meetings among distributed partners and participants. It provides the option to add and manage detailed agendas,

²¹ Note the use of the “DCH-RP.” namespace in the name of the DCH-RP Wiki landing page.

²² For example, see <https://documents.egi.eu/document/390>

²³ See also <https://documents.egi.eu/public>ListTopics> for details

assign session and track conveners, store meeting material and contributions, and participants to register for attendance.

EGI.eu provides the meeting planner using the Open Source tool Indico²⁴, integrated with the EGI SSO service. A top-level entry for the DCH-RP project available²⁵ and can be extended into further sub-categories as required.

8.1.6 Discussion forum

EGI.eu is providing a discussion forum²⁶ for a very wide variety of topics. This is a very young element of EGI.eu's collaboration services portfolio. It is based on the Open Source tool phpBB.

8.1.7 Blog

EGI provides a blogging facility that is integrated with the EGI SSO service. It is an integral part of the EGI web presence and powered by the Open Source tool Pebble²⁷. It features many elements of blogging; key features are editorial functions (e.g. when to publish blogs entries, editorial and publishing roles, commenting management, article excerpts, tagging, tag clouds).

Blogs are a meaningful outreach and connecting tool in the social media sphere. It facilitates feedback from targeted user communities, though in-depth discussions should be held using a discussion forum (see above).

²⁴ <http://indico-software.org/>

²⁵ <http://indico.evi.eu/indico/> and <http://indico.evi.eu/indico/categoryDisplay.py?categoryId=87>

²⁶ <https://forum.evi.eu/>

²⁷ <http://pebble.sourceforge.net/>