**DELIVERABLE**

|  |  |
| --- | --- |
| **Project Acronym:** | **DCH-RP** |
| **Grant Agreement number:** | **312274** |
| **Project Title:** | **Digital Cultural Heritage Roadmap for Preservation – Open Science Infrastructure for DCH in 2020** |

**D5.3 – Report on first Proof of Concept**

**Revision: Draft**

**Authors:**

**Michel Drescher (EGI.eu)**

**Eva Toller (RA)**

**Rosette Vandenbroucke (Belnet)**

**……**

|  |  |  |
| --- | --- | --- |
| **Project co-funded by the European Commission within the ICT Policy Support Programme** | | |
| **Dissemination Level** | | |
| **P** | **Public** | **P** |
| **C** | **Confidential, only for members of the consortium and the Commission Services** |  |

**Revision History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Revision** | **Date** | **Author** | **Organisation** | **Description** |
| Draft | 9-9-13 | M Drescher | EGI.eu | Initial skeleton |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Statement of originality:**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1 EXECUTIVE SUMMARY 4

2 Introduction 5

2.1 Objectives of the deliverable 5

2.2 Structure of the document 5

3 From Proofs of Concept To Scenarios 7

4 Reviewing the Management methodolgy 8

5 Scenario theme: “Organisational challenges” 11

5.1 Scenario 1.1 Using specialised research tools 11

5.2 Scenario 1.2 Integrating a new tool into an existing institutional infrastructure 11

5.3 Scenario 1.3 Selecting a digital preservation solution in the case of an institution with only voluntary IT support 11

5.4 Scenario 1.4 Preservation from a consortium of collections on the cloud 11

5.5 Scenario 1.5 Preserving a 3D visualisation 11

5.6 Scenario 1.6 Archived data retrieving 11

6 Scenario Theme: “End user concerns” 12

6.1 Scenario 2.1 Researcher discovers a historical database 12

6.2 Scenario 2.2 Research and select a tool serving a specific purpose 12

6.3 Scenario 2.3 Accessing digitised content from schools(?) 12

6.4 Scenario 2.4 Gain access to archived websites 12

7 Conclusion 13

ANNEx 1: Technical Interpretation of the Scenarios 14

ANNEx 2: Scenario report template 15

A2.1 Document metadata 15

A2.2 Executive summary and grading 15

A2.3 Proof of Concept report 15

A2.4 Annexes 15

ANNEx 3: Preservation of the WP5 Scrum backlog 17

# EXECUTIVE SUMMARY

# 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.

Work Package 5, which is responsible for producing and delivering this document, has been charged with coordinating and conducting Proofs of Concept (PoCs) that will inform Work Package 3 with the outcomes of these PoCs so that informed decisions can be made towards further evolution of the DCH roadmap to preservation within Work Package 3.

Deliverable D5.1 [REF], made available in January 2013 to the general public, informs about the planning around the Proofs of Concept; as such it is a prospective document in nature and provides information on the work package management methodology (using SCRUM), the intentions and objectives of the participating national partners, and the general planning and coordination with other Work Packages in the DCH-RP project.

This document, Deliverable D5.3, is retrospective in nature by reporting on the conducted Proofs of Concepts in the past seven months (from 18 February 2013 to 30 September 2013) and prospective in indicating next steps and lessons learnt that would influence the second PoCs phase that will begin immediately after publication of this document.

## Objectives of the deliverable

This deliverable formally informs the reader about the outcomes of the conducted Proofs of Concept. The primary recipients of this deliverable are the members of Work Package 3, which is charged with maintaining and updating the DCH roadmap based on information contained in this document.

Throughout the operative timespan in the first PoC phase, the work package members have agreed to produce a reporting template, which will be used to inform the roadmap maintainers in Work Package 3. For each “scenario” (see below) that was examined by participating national institutes, a scenario report based on this template was produced and made permanently available in the EGI Document DB. Since the leaders of Work Package 3 and Work Package 4 are closely involved in the activities in Work Package 5 as the “Product Owners” of the PoC activities, this approach was implicitly approved as the means of information transfer to Work Package 3.

Therefore, these scenario reports comprise the core of WP 4’s feedback to WP 3. To avoid duplication of work, this deliverable incorporates these reports by references provided later in this document, thus fulfilling the contractual obligation formulated in the DCH-RP DoW.

## Structure of the document

Reflecting the decisions made during the first phase of Proofs of Concept, this document is structured as follows.

Section 3 summarises how the WP5 participants have started to design the first Proofs of Concept phase; beginning with a cross Work Package discussion with WP3 and WP4, a first list of scenarios that were provided by the DCH roadmap maintainers and the coordination of the individual scenario tests across member institutes in WP5.

By adopting the Scrum methodology of agile project management for WP5, the team regularly reviewed the last sprint in a “retrospection” session in the sprint planning meetings. Consequently, section 4 provides a retrospective analysis of the experience of applying Scrum to a physically distributed team of project members.

Section 5 summarises the activities and results of the individual Proofs of Concept with the common theme “Organisational Challenges”. The concrete and final scenario reports are incorporated by reference in this document and were individually forwarded to Work Package 3 for further consideration in a roadmap review at a later date.

Similarly, section 6 covers the Proofs of Concept around the common theme “End User Concerns”.

Section 7 concludes this deliverable summarising the lessons learnt across all member institutes, Proofs of Concepts. It will highlight recommendations to Work Package 3, and indicate the next steps during the second, final Proofs of Concept phase until the conclusion of the DCH-RP project.

Annex 1 provides the discussion document that includes the technical interpretation of the scenarios indicated by Work Package 3, which led to the scenarios that were examined in this first phase of Proofs of Concept.

Annex 2 includes the scenario report template that was used by team members to report back to Work Package 3.

Annex 3 preserves a snapshot of the contents of the online maintained Scrum backlog for long-term reference.

# From Proofs of Concept To Scenarios

<<Describe here how we arrived at the scenario overview that drove WP5’s decision to work towards scenarios. Discuss, in greater detail, the lessons learnt as presented in the slides at the Plenary.>>

# Reviewing the Management methodolgy

As extensively described in D5.1[[1]](#footnote-1) WP5 has adopted the Scrum methodology for managing and running the Proofs of Concept. The methodology, used tools and facilities are also described in D5.1 and will not be repeated here.

A key role in managing a Scrum Product Team working on a specified product (here: the Proofs of Concept and the resulting reports) lies in planning and conducting the so called “Sprints” – focused phases of productivity that subsequently finish with a working end result.

The planning of Scrum sprints revolves around one key tool, the Sprint history. Usually recorded in a spreadsheet, the Sprint history captures key indicators about the Product Team’s progress on finishing tasks that are recorded in the backlog. Annex 3 provides a complete listing of the sprint history of the first PoC phase.

Of these indicators, four are used not only for historic recording of data, but also to *project* the future progress of the team based on the historic data. Hence the Sprint history becomes an intuitive and effective tool to **manage expectations** towards the Product Owner, i.e. the one who oversees and steers toe progress of the team. These key indicators are:

1. **Performance**The performance is the aggregation of the points of all tasks that were finished by the team, and accepted by the Product Owner. If a team has not finished any task in a sprint (e.g. Sprint 9), or the Product Owner did not accept any of the finished tasks, then the team performance for that sprint is recorded as 0 (zero).
2. **Velocity**  
   The team’s velocity is expressed as an averaging function over the team’s productivity. Typically, the velocity is measured as the average productivity over the team’s last three sprints – and so does this team. Ideally, Productivity and Velocity gradually close in on each other with the team’s gain in experience in the Scrum methodology. The *current* velocity is used as the expected performance of the team for all future sprints.
3. **Remaining**  
   The remaining points are calculated as the difference between the total points at the end of a sprint, and the sum of all points of all tasks accomplished in the current and previous sprints. The remaining points are then the basis of projecting into the future (using the team’s velocity) to calculate how many more sprints are needed to accomplish all work encoded in the task points.
4. **Total points**The total points are simply the aggregation of all points of all tasks in the backlog.

The following figure provides an overview of the historic development of these four key indicators during the first phase of Proofs of Concepts.

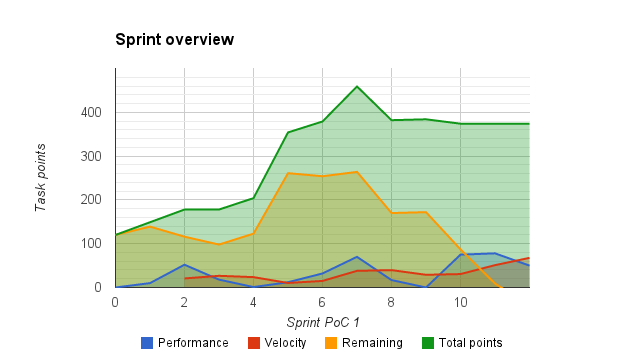


Figure 1: A graphical overview of WP5's sprint history

The graph shows some typical mechanics of how Scrum teams work.

Firstly, the **total points** fluctuate, in this case quite drastically. This is normal for a Scrum Team as WP5 since this methodology was new to all team members, so a learning phase of about three to five Sprints indicates the team getting used to a number of processes within the Scrum methodology. A key process is the allocation of points to tasks, where the task represents a unit of work, and the points reflect the team’s expected amount of effort necessary to accomplish this task. With the learning curves increases the tea’s accuracy in assigning points – in this case the team had to learn to be more realistic (i.e. reduce expectations and factor in unexpected problems) and allocate more points to the task than initially anticipated.

The steep increase in total points is, however, only partially explained by the initial learning phase. As often seen in Scrum projects, the initial phase is followed by a phase that is best described with “explosion of creativity” in team members, adding many tasks that members feel the need to accomplish to reach the overall goal.

As the team progressed, the projection feature of the Scrum methodology allowed the team to *manage itself* in that the realization began to appear that not everything that the team thought it needed to accomplish in fact should be accomplished during the first phase of the Proofs of Concept. This started to happen in Sprint 7 and onwards, illustrated by the fact that the numbers of total points, remaining points and performance do not add up: The difference is caused by a number of tasks (with story points) that were moved to the second phase of Proofs of concept, as detailed in Annex 3.

The **remaining points** follow the dynamics of total points and performance, except for when the team decided to move tasks to the second PoC phase (see above). This is attributed to the main purpose of the remaining points indicator in Scrum: In the projection for future sprints, the remaining points graph provides the team with feedback as to where they are standing in the middle of the project. It also allows the Product Owner as well as the Team to make an informed decision for project planning. In a situation like DCH-RP, the number of sprints is fixed as determined by the DoW, therefore the project cannot extend the duration to accomplish all planned tasks. Instead, a decision must be taken, a prioritisation over all remaining tasks in the project. The result is that a number of tasks are cut or in this case, moved to the second PoC phase. This project management situation is perfectly reflected in this graph and explained earlier.

The **velocity** indicator works as expected. Its averaging effect makes it a good progression forecast tool for project management decision. As indicated, the velocity is not available for the first two sprints – this team’s velocity is calculated as the average over the last three sprints.

The **performance** graph shows significant peaks over time, with stark troughs in between. Usually, the performance fluctuates in the beginning, and closes in on a steady state over the time of a project. This behavior, however, is typical only for Scrum teams that wok exactly as the Scrum methodology suggests, i.e. a group of participants that are physically located very close to each other – ideally in the same office or even room. *This* product team, however, is quite different from the ideal Scrum setup:

* The team members are distributed over all Europe
* The team members are not working 100% of their time on this project
* The team members act as proxies to local colleagues who often conduct the work.
* Those local colleagues are not included in the Scrum methodology, nor do they work exclusively in this project.

While the troughs can be broadly explained by the vacation periods around Easter and Summer 2013, the main reason of the fluctuations are due to the setup of the team detailed above.

Nonetheless, the team achieved what it had set out to achieve, and the Scrum methodology has proved a useful tool to manage a distributed team of devoted participants. It does not come, therefore, as a surprise that the team decided to continue following the Scrum methodology for the second Proof of Concept phase. However, the means of communications need to be improved. The team expressed its expectations that, with better means of communications during the Scrum sprints, the overall effectiveness of the team could be greatly improved.

# Scenario theme: “Organisational challenges”

<<Group reports on the “organizational challenges” theme of scenarios>>

## Scenario 1.1 Using specialised research tools

## Scenario 1.2 Integrating a new tool into an existing institutional infrastructure

## Scenario 1.3 Selecting a digital preservation solution in the case of an institution with only voluntary IT support

## Scenario 1.4 Preservation from a consortium of collections on the cloud

## Scenario 1.5 Preserving a 3D visualisation

## Scenario 1.6 Archived data retrieving

# Scenario Theme: “End user concerns”

## Scenario 2.1 Researcher discovers a historical database

## Scenario 2.2 Research and select a tool serving a specific purpose

## Scenario 2.3 Accessing digitised content from schools(?)

## Scenario 2.4 Gain access to archived websites

# Conclusion

# ANNEx 1: Technical Interpretation of the Scenarios

<<Add in here the document of the scenarios (see section 2) and their interpretation as a document circulated by Borje and Rosette>>

# ANNEx 2: Scenario report template

During implementing the Proofs of Concept the members of WP4 decided to report the results of the conducted activities in individual reports, rather collecting them in one deliverable. This approach has several benefits, in that it allows a more individual reporting, focusing on the actual work undertaken, and the respective results. It allows the roadmap evolution activity in WP3 to ingest the results as they see fit. Further, it reflects the more particular nature in the Proofs of Concepts that were not connected to each other, and conducted in a more national fashion (except for Belgian and Italian partners).

This section incorporates the reporting template in a condensed form, highlighting the core benefits and key elements. The full report template is maintained in the persistent document storage (<https://documents.egi.eu/document/1892>).

## A2.1 Document metadata

The first part of the report template contains all necessary metadata about the information contained in the document. This includes enumerating the scenarios and any tools and services that were used in the Proof of Concept:

* Covered scenario, and tools
* Authors
* Revision of the document (including final)
* Revision history
* Dissemination level
* Persistent storage location (in the footer of the document)

## A2.2 Executive summary and grading

The first content section provides an executive summary, along with a tabular overview of the grading of a number of aspects that apply to the described Proof of Concept. The template does not impose any aspects to be graded; however each aspect that is included in the summary must be sufficiently explained in the annex of the report. Aspects capture anything that is important for the Proof of Concept and subsequent assessment in roadmap discussions. Aspects may range from functionality (e.g. “Support of Dublin Core metadata”) to non-functional requirements, such as ease of installation.

The executive part of the report also includes specific recommendations the authors wish to highlight to the roadmap evolution team.

## A2.3 Proof of Concept report

The bulk of the report contains the details of the conducted Proof of Concept. While the template suggests separating the description of the Scenario from the tools tested within the scenario(s), it is intentionally left to the authors how to structure this part of the document: Form follows function, for as long as the key outcomes are reflected accordingly in the executive summary section.

## A2.4 Annexes

The template provides only one annex, but does not limit the authors to add more annexes where required. The template’s sole required Annex lists all aspects of the involved tools and services that were examined throughout the Proof of Concept.

The description of each aspect is designed to be complete and self-referential within the document, in that it defines:

* Aspect name
* Description
* Grades “n/a”, 1, 2, 3, 4 and 5
* Semantics of each grade

This way, WP4 is able to collect a set of specific aspects that were used and tested during the Proofs of Concept. These may overlap, but the process of reconciliation is expected to be negligible. The resulting set of aspects then can be used for future reference in further Proofs of Concept.

# ANNEx 3: Preservation of the WP5 Scrum backlog

The Scrum backlog is kept and maintained in an online collaborative document[[2]](#footnote-2) stored in Google’s Cloud storage “Google Drive”. It is accessible and editable by all WP5 members. This annex is preserving the raw data from the first PoC phase that formed the basis of the day-to-day work in the last months.

Table 1: All "user stories" (tasks) that were accomplished or rejected during the first PoC phase.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Epic** | **ID** | **Priority** | **Task** | **Points** | **Status** |
| 1 | 1 | 3 | 5 | Find and confirm up to 4 Italian CH partners for PoCs 1 & 2 | 0 | Accepted |
| 1 | 1 | 4 | 5 | Find and confirm up to 4 Polish CH partners for PoCs 1 & 2 | 0 | Accepted |
| 1 | 2 | 7 | 4 | Find out EGI NGI International Liaison & deputy for Estonia | 1 | Accepted |
| 1 | 2 | 8 | 4 | Find out EGI NGI International Liaison deputy for Poland | 1 | Accepted |
| 1 | 2 | 9 | 4 | Find out EGI NGI International Liaison & deputy for Sweden | 1 | Accepted |
| 1 | 4 | 10 |  | Confirm BEgrid support through its NIL | 0 | Accepted |
| 1 |  | 14 |  | Confirm BEgrid Grid Resource Providers | 0 | Accepted |
| 1 |  | 20 | 4 | Confirm BEgrid Cloud Resource Providers | 3 | Accepted |
| 1 |  | 40 | 5 | Set up WP5 group and mailing list for DCH-RP | 2 | Accepted |
| 1 |  | 41 | 5 | Find time slot for Sprint 1 review meeting | 1 | Accepted |
| 1 |  | 42 | 5 | Find timeslot for Sprint 2 planning meeting | 1 | Accepted |
| 2 | 1 | 2 | 5 | Find and confirm up to 4 Hungarian CH partners for PoCs 1 & 2 | 1 | Accepted |
| 2 | 1 | 5 | 5 | Find and confirm up to 4 Swedish CH partners for PoCs 1 & 2 | 0 | Accepted |
| 2 | 2 | 6 |  | Find out EGI NGI International Liaison & deputy for Belgium | 1 | Accepted |
| 2 | 2 | 11 | 4 | Confirm EEnet support through its NIL | 2 | Accepted |
| 2 | 2 | 12 | 4 | Confirm MGKK support through its NIL | 2 | Accepted |
| 2 | 2 | 13 | 4 | Confirm IGI support through its NIL | 2 | Accepted |
| 2 | 2 | 16 | 4 | Confirm MGKK Grid Resource Providers | 3 | Accepted |
| 2 | 2 | 17 | 4 | Confirm IGI Grid Resource Providers | 3 | Accepted |
| 2 | 2 | 18 | 4 | Confirm PL-Grid Grid Resource Providers | 3 | Accepted |
| 2 | 2 | 19 | 4 | Confirm SweGrid Grid Resource Providers | 3 | Accepted |
| 2 | 2 | 21 | 4 | Confirm EEnet Cloud Resource Providers | 3 | Accepted |
| 2 | 2 | 22 | 4 | Confirm MGKK Cloud Resource Providers | 3 | Accepted |
| 2 | 2 | 23 | 4 | Confirm IGI Cloud Resource Providers | 3 | Accepted |
| 2 | 2 | 24 | 4 | Confirm PL-Grid Cloud Resource Providers | 3 | Accepted |
| 2 |  | 30 | 2 | Decide on Cooperation Agreement for Italy | 0 | Accepted |
| 2 |  | 34 | 5 | Provide a link to the installed e-cultural Science gateway on the wiki | 1 | Accepted |
| 2 |  | 43 | 5 | Document what needs to be documented where | 3 | Accepted |
| 2 |  | 44 | 5 | Understand and document Scenario 1.1 | 5 | Accepted |
| 2 |  | 45 | 5 | Understand and document Scenario 1.2 | 5 | Accepted |
| 2 |  | 46 | 5 | Understand and document Scenario 1.3 | 5 | Accepted |
| 2 |  | 64 | 5 | Provide your Easter holiday details to Michel (covering sprint 3 & 4) | 1 | Accepted |
| 3 | 2 | 25 | 4 | Confirm SweGrid Cloud Resource Providers | 3 | Accepted |
| 3 |  | 35 | 5 | Provide a link to the e-cultural Science gateway documentation on the wiki | 1 | Accepted |
| 3 | 3 | 49 | 5 | Decide on test data for scenario 1.1 | 2 | Accepted |
| 3 | 5 | 51 | 5 | Decide on test data for scenario 1.3 | 2 | Accepted |
| 3 | 5 | 56 | 4 | Describe testing procedures for scenario 1.3 | 3 | Accepted |
| 3 | 9 | 66 | 5 | Decide on test data for scenario 2.2 | 1 | Accepted |
| 3 | 11 | 68 | 5 | Decide on test data for scenario 2.4 | 1 | Accepted |
| 3 | 9 | 70 | 4 | Describe testing procedures for scenario 2.2 | 3 | Accepted |
| 3 | 11 | 72 | 4 | Describe testing procedures for scenario 2.4 | 2 | Accepted |
| 4 |  | 33 | 5 | Update https://documents.egi.eu/document/1602 with Epic 1 resilts | 1 | Accepted |
| 5 | 3 | 59 | 2 | Identify required e-Infrastructure resources for scenario 1.1 | 1 | Accepted |
| 5 | 4 | 60 | 3 | Identify required e-Infrastructure resources for scenario 1.2 | 1 | Accepted |
| 5 | 6 | 62 | 3 | Identify required e-Infrastructure resources for scenario 1.4 | 1 | Accepted |
| 5 | 3 | 77 | 3 | Identify required local resources for scenario 1.1 | 1 | Accepted |
| 5 | 3 | 78 | 3 | Identify required local resources for scenario 1.2 | 1 | Accepted |
| 5 | 3 | 80 | 3 | Identify required local resources for scenario 1.4 | 1 | Accepted |
| 5 | 5 | 81 | 3 | Testing de-identification with ROND (Riksarkivet Open Data) Scenario 1.3 | 1 | Accepted |
| 5 | 5 | 83 | 3 | Obtaining permission to use test data outside the Preservation Net, scenarios 1.3, 2.2, 2.4 | 1 | Accepted |
| 5 |  | 87 | 3 | B PoC2 Ask membership of the VO DCH-RP and Indicate | 3 | Accepted |
| 5 |  | 88 | 4 | B PoC 2 Next Step in mastering the eCSG - attend Webinar 15/5 | 1 | Accepted |
| 6 | 1 | 1 | 5 | Find and confirm up to 4 Estonian CH partners for PoCs 1 & 2 | 1 | Accepted |
| 6 |  | 26 | 2 | Confirm Cooperation Agreement document & contents | 0 | Accepted |
| 6 |  | 37 | 5 | Organise a training session on the e-CSG | 2 | Accepted |
| 6 | 4 | 50 | 5 | Decide on test data for scenario 1.2 | 1 | Accepted |
| 6 | 6 | 52 | 5 | Decide on test data for scenario 1.4 | 5 | Accepted |
| 6 | 3 | 54 | 4 | Describe testing procedures for scenario 1.1 | 5 | Accepted |
| 6 | 4 | 55 | 4 | Describe testing procedures for scenario 1.2 | 5 | Accepted |
| 6 | 6 | 57 | 4 | Describe testing procedures for scenario 1.4 | 1 | Accepted |
| 6 | 3 | 79 | 3 | Identify required local resources for scenario 1.3 | 3 | Accepted |
| 6 | 9 | 84 | 3 | Identify required local resources for scenario 2.2 | 3 | Accepted |
| 6 | 9 | 85 | 3 | Selecting specific images for test data, scenario 2.2 | 3 | Accepted |
| 6 | 11 | 86 | 3 | Identify required local resources for scenario 2.4 | 3 | Accepted |
| 7 |  | 47 | 4 | Understand and document Scenario 1.4 | 5 | Accepted |
| 7 |  | 48 | 3 | Understand and document Scenario 1.5 | 5 | Accepted |
| 7 | 5 | 61 | 3 | Identify required e-Infrastructure resources for scenario 1.3 | 5 | Accepted |
| 7 | 9 | 74 | 3 | Identify required e-Infrastructure resources for scenario 2.2 | 3 | Accepted |
| 7 | 11 | 76 | 3 | Identify required e-Infrastructure resources for scenario 2.4 | 3 | Accepted |
| 7 |  | 89 | 4 | B PoC2 Exploit the possibility to have eCSG working with dCache | 13 | Accepted |
| 7 |  | 90 | 4 | Look at the integration of metada | 20 | Accepted |
| 7 |  | 91 | 3 | B PoC2 if dcache works with eCSG add the Belgian gridstorage to the VO DCH-RP | 5 | Accepted |
| 7 |  | 92 | 3 | B PoC 2 if 86 is positive, copy data from KIK to the Belgian gridstorage | 5 | Accepted |
| 7 | 9 | 109 | 3 | S PoC 1 Decide target formats for conversion of images | 1 | Accepted |
| 7 | 9 | 110 | 3 | S PoC 1 Test the AVS Document Converter | 3 | Accepted |
| 7 | 5 | 119 |  | S PoC 1 Metadata: check the status of the Swedish eARD project for the chosen file formats | 2 | Accepted |
| 8 |  | 111 |  | B PoC 2 Italian partners to become member of the VO DCH-RP | 3 | Accepted |
| 8 |  | 112 |  | B PoC 2 Italian partners to check readiness to use the Italian e-CSG | 3 | Accepted |
| 8 |  | 113 | 2 | B PoC 2 Italian partners to make the choice of data to go to the gridstore | 5 | Accepted |
| 8 |  | 114 |  | B PoC 2 Italian partners to check data for transfer to the Italian Grid storage | 3 | Accepted |
| 8 |  | 138 |  | B PoC 1 testing the scoremodel | 3 | Accepted |
| 10 | 11 | 120 |  | S PoC 1 Test WARC Tools | 40 | Accepted |
| 10 | 5 | 121 |  | S PoC 1 Test Archivist's Toolkit | 5 | Accepted |
| 10 | 11 | 122 |  | S PoC 1 Test Web Curator Tool | 3 | Accepted |
| 10 | 11 | 123 |  | S PoC 1 Test SWAT | 5 | Accepted |
| 10 | 11 | 125 |  | S PoC 1 Obtain a test person (like the one described in scenario 2.4) | 1 | Accepted |
| 10 | 9 | 126 |  | S PoC 1 Test Universal Document Converter | 3 | Accepted |
| 10 | 5 | 127 |  | S PoC 1 Test DSpace | 3 | Accepted |
| 10 | 9 | 128 |  | S PoC 1 Test A-PDF DjVu to PDF | 3 | Accepted |
| 10 | 11 | 129 |  | S PoC 1 Test HTTRACK | 3 | Accepted |
| 10 | 5 | 133 |  | S PoC 1 Test XENA | 3 | Accepted |
| 10 | 9 | 136 |  | S PoC 1 Test AVS Image Converter | 3 | Accepted |
| 10 | 11 | 137 |  | S PoC 1 Test Heritrix | 3 | Accepted |
| 11 | 2 | 15 | 4 | Confirm EEnet Grid Resource Providers | 3 | Accepted |
| 11 |  | 93 | 2 | B PoC2 Define the access methods to the grid storage (depends on #139) | 5 | Accepted |
| 11 |  | 94 | 3 | B PoC 2 Define the access measurement tools (depends on #139) | 20 | Accepted |
| 11 |  | 115 | 1 | B PoC 2 Italian partners copy their data to the Italian grid storage | 20 | Accepted |
| 11 |  | 139 | 1 | B PoC 1 Belgian partners to copy their data into the Italian Grid storage | 20 | Accepted |
| 11 |  | 95 | 3 | B POC2 Do the data access tests | 5 | Accepted |
| 12 |  | 104 | 1 | B PoC 1 report the PoC | 3 | In Progress |
| 12 |  | 140 | 1 | B PoC2 Report the PoC | 3 | In Progress |
| 11 | 11 | 130 |  | S PoC 1 Let the test person find the archive Linnéjubileet, open it and evalutate how easy it was to find and use this archive | 2 | Accepted |
| 11 | 11 | 131 |  | S PoC 1 Compare how the access to Linnéjubileet works when you use the original one (www.linne2007.se) and when you use the format/tools chosen for scenario 2.4. (This test must be omitted if the original site ceases to be accessible) | 3 | Accepted |
| 12 |  | 134 |  | S PoC 1 Report the PoC to WP3 | 3 | Accepted |
|  |  | ~~124~~ |  | ~~S PoC 1 Register the web site archive for Linnéjubileet in the archive information system NAD (“Nationella arkivdatabasen”, the National Archive Database)~~ |  | ~~Accepted~~ |
|  |  | ~~36~~ | ~~5~~ | ~~Provide link to documentation on dARCEO on the Wiki~~ |  | ~~Rejected~~ |
|  |  | ~~38~~ | ~~5~~ | ~~Start the registry of services documentation in the Wiki~~ |  | ~~Rejected~~ |
|  |  | ~~39~~ | ~~5~~ | ~~Add a first set of tools and services to the registry~~ |  | ~~Rejected~~ |
|  | ~~7~~ | ~~53~~ | ~~5~~ | ~~Decide on test data for scenario 1.5~~ |  |  |
|  | ~~7~~ | ~~58~~ | ~~4~~ | ~~Describe testing procedures for scenario 1.5~~ |  |  |
|  | ~~7~~ | ~~63~~ | ~~3~~ | ~~Identify required e-Infrastructure resources for scenario 1.5~~ |  |  |
|  | ~~8~~ | ~~65~~ | ~~5~~ | ~~Decide on test data for scenario 2.1~~ |  |  |
|  | ~~10~~ | ~~67~~ | ~~5~~ | ~~Decide on test data for scenario 2.3~~ |  |  |
|  | ~~8~~ | ~~69~~ | ~~4~~ | ~~Describe testing procedures for scenario 2.1~~ |  |  |
|  | ~~10~~ | ~~71~~ | ~~4~~ | ~~Describe testing procedures for scenario 2.3~~ |  |  |
|  | ~~8~~ | ~~73~~ | ~~3~~ | ~~Identify required e-Infrastructure resources for scenario 2.1~~ |  |  |
|  | ~~10~~ | ~~75~~ | ~~3~~ | ~~Identify required e-Infrastructure resources for scenario 2.3~~ |  |  |
|  | ~~3~~ | ~~82~~ | ~~3~~ | ~~Identify required local resources for scenario 1.5~~ |  |  |
|  |  | ~~135~~ |  | ~~S PoC 1 Report the PoC to WP3~~ |  |  |

Table 2: Some tasks were decided to be tackled in the second PoC phase.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Epic** | **ID** | **Priority** | **Task** | **Points** | **Status** |
| 13 |  | 27 | 2 | Decide on Cooperation Agreement for Belgium | 60 |  |
| 13 |  | 28 | 2 | Decide on Cooperation Agreement for Estonia |  |  |
| 13 |  | 29 | 2 | Decide on Cooperation Agreement for Hungary |  |  |
| 13 |  | 31 | 2 | Decide on Cooperation Agreement for Poland |  |  |
| 13 |  | 32 | 2 | Decide on Cooperation Agreement for Sweden |  |  |
| 13 |  | 98 | 2 | B PoC1 definition of documentation of the work | 13 | In Progress |
| 13 |  | 99 | 4 | B-PoC1 choice of audit tool (Drambora or ...) with partners | 20 | In Progress |
| 13 |  | 101 | 4 | B PoC 1 choice of part of the tool if necessary | 20 | In Progress |
| 13 | 5 | 107 | 3 | S PoC 2 Identify Swedish "cloud archive providers" | 3 | In Progress |
| 13 | 5 | 108 | 3 | S PoC 2 Compile a list of requirements for "cloud archive providers" | 3 | In Progress |
| 14 | 5 | 116 |  | S PoC 2 Construct a questionnaire and send it to DCH-RP WP5 and other interested parties for review | 3 |  |
| 14 | 5 | 117 | 3 | S PoC 2 Update questionnaire and send it to the "cloud archive providers" | 3 |  |
| 19 | 5 | 118 | 3 | S PoC 2 Evaluate the answers to the questionnaire and determine if a commercial "cloud archive provider" is suitable for a small institution like that in scenario 1.3 | 5 |  |
|  | 5 | 106 | 3 | S PoC 1 Test and evaluate the ARC Graphical Client (for SweGrid/SweStore) | 5 |  |
| 13 |  | 96 | 1 | B PoC2 Exploit the possibility to install eCSG on BEgrid | 20 | In Progress |
| 13 |  | 97 | 1 | B PoC2 Exploit the data access tests on the Belgian eCSG | 5 | In Progress |
| 13 |  | 100 | 2 | B PoC1 definition of the local preserved data on which the audit is done | 3 | In Progress |
| 13 |  | 102 | 2 | B PoC1 definition of the audit procedures | 13 | In Progress |
| 13 |  | 103 | 1 | B PoC 1 execution of the tasks | 13 | In Progress |
| 13 | 5 | 105 | 3 | S PoC1 Obtaining permission to use test data outside Riksarkivet, scenarios 1.3, 2.2, 2.4 | 3 | In Progress |
| 13 | 5 | 132 |  | S PoC 1 Test Fedora | 3 |  |

Table 3: Each sprint is maintained with start and end dates, and key indicators for planning purposes.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Start** | **End** | **duration [d]** | **Sprint** | **Theme** | **Performance** | **Velocity** | **Remaining** | **Total points** |
| **PoC 1** | 18 Feb 2013 | 3 Mar 2013 | 14 | 0 | Setting up Scrum | 0 | n/a | 120 | 120 |
| 4 Mar 2013 | 17 Mar 2013 | 14 | 1 | Connecting with e-Infrastructures | 10 | n/a | 139 | 149 |
| 18 Mar 2013 | 4 Apr 2013 | 18 | 2 | Understanding scenarios | 52 | 20.67 | 116 | 178 |
| 4 Apr 2013 | 21 Apr 2013 | 18 | 3 | Working on test data and procedures | 18 | 26.67 | 98 | 178 |
| 22 Apr 2013 | 6 May 2013 | 15 | 4 | Expanding work on PoC scenarios | 1 | 23.67 | 123 | 204 |
| 6 May 2013 | 23 May 2013 | 18 | 5 | Gather our first concrete results | 12 | 10.33 | 261 | 354 |
| 23 May 2013 | 10 Jun 2013 | 19 | 6 |  | 32 | 15.00 | 254 | 379 |
| 10 Jun 2013 | 24 Jun 2013 | 15 | 7 |  | 70 | 38.00 | 264 | 459 |
| 24 Jun 2013 | 15 Jul 2013 | 22 | 8 | Vacation, vacation, vacation ;) | 17 | 39.67 | 170 | 382 |
| 15 Jul 2013 | 29 Jul 2013 | 15 | 9 | Vacation, vacation, vacation ;) | 0 | 29.00 | 172 | 384 |
| 29 Jul 2013 | 13 Aug 2013 | 16 | 10 | Finish tests that are in progress | 75 | 30.67 | 87 | 374 |
| 13 Aug 2013 | 9 Sep 2013 | 28 | 11 | Finish results and draft D5.3 | 78 | 51.00 | 9 | 374 |
| 9 Sep 2013 | 30 Sep 2013 | 22 | 12 | Write and deliver D5.3 | 50 | 67.67 | -41 | 374 |

Table 4: Epics are a means of grouping tasks together in a meaningful way.

|  |  |  |
| --- | --- | --- |
| Epic ID | Title | Description |
| 1 | Establish CH institute cooperation | Find CH institutes that are willing to contribute their data to the PoCs, as well as assume the role of users of the infrastructures set up in the PoCs. |
| 2 | Establish EGI resources | EGI is offering Grid and Cloud resources for the DCH-RP project. Establish and confirm these resource on a national & local coordinated way. |
| 3 | Scenario 1.1 Using specialised research tools | A major memory institution in France which has its own development team is gradually implementing a solution for digital preservation. It is using local in-house storage. The institution participates in projects which aggregate content to Europeana and regularly uses social media channels to engage with the wider public. Thus, the access to its digital collections is either possible through the institutional website, or resource discovery is made via specialised portals and social media which in fact redirect the users to the institutional webserver. Recently, it has happened several times that researchers ask to use specialised document analysis tools that are available through an e-Infrastructure. This raises issues of sharing content outside the institutional storage and preservation facilities on the cloud used by the eInfrastucture, or the use of ‘external’ tools for processing locally stored documents. Both options raise concerns, and for the time being there is no good solution for the end users. |
| 4 | Scenario 1.2 - Integrating a new tool into an existing institutional infrastructure | A major memory institution in Germany had already developed its own preservation infrastructure. A new research project is asking for a newly developed software tool that would save time on checking file formats. However, the integration of this tool with the existing preservation solution cannot compromise any essential preservation features implemented in the local preservation system. The requirement is to analyse the difference that using the new tool will make and how to embed it with other components already in place; or how to run the new tool from a cloud-based provider and integrate this service with the existing preservation solution. |
| 5 | Scenario 1.3 - Selecting a digital preservation solution in the case of an institution with only voluntary IT support | A little museum in Malta has a historical library and a digitised personal archive collection. The museum has staff of only 9 and only voluntary IT support. The director of the museum is aware of the need to organise digital preservation for the digitised documents, but is not sure how to do it. He receives periodically offers for long-term storage of digital content, but finds it difficult to select or to make a decision. He has practically no IT competence to rely on for decision-making, but is convinced that the decision should be forward-looking and accommodate the needs of the museum for the next 5 years. |
| 6 | Scenario 1.4 - Preservation from a consortium of collections on the cloud | A specialised consortium of several institutions working on a complete digital repository of the works of a modern digital artist who worked and exhibited in 15 different countries has to resolve the issue of preservation of objects that are stored in different location. The works of the digital artist include a variety of digital formats as well as especially developed software tools. The curator of the collection has to identify a cost efficient solution which would also be suitable to store the complex objects in the collection. An additional difficulty is that the copyrights on the objects differ in the countries of origin of the objects. |
| 7 | Scenario 1.5 - Preserving a 3D visualisation | A research lab in the UK is collaborating with an archaeological site in Italy to create a 3D visualisation of an ancient building. The visualisation is used as scientific documentation. Both institutions have to agree who will take care for the preservation in usable state of the model. There is also an issue of interoperability of the model with a free visualisation tool which can be used to show the model on a web site which is resolved producing a lower quality visualisation in an additional format. There is an ongoing discussion whether it also needs to be preserved and by whom. |
| 8 | Scenario 1.6 - Archived data retrieving | Estonian memory institution (Conservation Centre Kanut) which digitises different content wants to make backup copy of files to another memory institutions tape library but needs proof that content is well preserved and it is possible to receive copy of files if needed. Therefore periodically (quarterly) will be carried out test data retrieving. |
| 9 | Scenario 2.1 - Researcher discovers a historical database | Researcher in history discovers a historical database resource presenting parish records. She would like to use the data, but she is also concerned to what extent these data could be trusted (authenticity, error rates introduced, errors caused by any transformations needed). |
| 10 | Scenario 2.2 - Research and select a tool serving a specific purpose | A university lecturer in art history wants to use a collection of digitised art images made 15 years ago. They are stored in a format he is not familiar with. Since there are about 200 images, the researcher is looking for tools which would convert them into a format he could easily use in batch mode. He is not sure how to identify a tool or a service which could do this. |
| 11 | Scenario 2.3 - Accessing digitised content from schools(?) | Secondary school students are making an assignment looking at historical maps of their village. They already paid a visit to the local museum but discovered some old digitised maps on the internet. |
| 12 | Scenario 2.4 - Gain access to archived websites | A history student interested in natural history discovers that Riksarkivet has archived the "Linnéjubilet" web site http://www.riksarkivet.se/default.aspx?id=23153 .He wonders how he can get access to it (the link www.linne2007.se obviously doesn't work anymore). |
| 13 | Scenario 3.1 - Proof of authenticity in distributed archiving | The Swedish National Archives takes 10 digitised images of records and ingests them into their national GRID where they undergo a migration cycle or some other processing and the SNA requires a proof of authenticity at the end of this. |
| 14 | Scenario 3.2 - Defining new services | A small art gallery looks for the grid infrastructure for storage services that could solve the preservation problems. For that is needed new services not yet defined. |
| 15 | Scenario 3.3 - Integrating new services into existing infrastructure | The IT manager of a local art gallery is preserving the digital content using grid X. He attends a workshop on digital preservation where he hears about a new tool for checking the integrity of digital objects. He needs to implement it on the grid-based archiving solution. |
| 16 | Scenario 3.1a - extending 3.1 with repository safeguarding policies | During this processing , the GRID provider ask for information about SNA:s requirements for safeguarding a trustworthy repository. The SNA has to describe the methodology and tools they are useing for validating their objectives and methods as well as their management of intrinsic theats and threats originating from the outside of the organisation. The purpose from the GRID providers point of view is to push the SNA to come up with trust critieria for the services it will get from the GRID. |

Table 5: WP4 also kept track of which partner examined which scenario in their Proofs of Concept.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | | **Partner** | | | | | |
| Theme | Scenario | BELSPO (Belgium) | EVKM (Estonia) | NIIFI (Hungary) | ICCU (Italy) | PSNC (Poland) | RA (Sweden) |
| Organisational challenges | Scenario 1.1 | X |  |  | X |  |  |
| Scenario 1.2 | X |  |  |  |  |  |
| Scenario 1.3 |  |  |  |  |  | X |
| Scenario 1.4 | X |  |  |  |  |  |
| Scenario 1.5 |  |  |  |  |  |  |
| Scenario 1.6 |  | X |  |  |  |  |
| End user concerns | Scenario 2.1 |  |  |  |  |  |  |
| Scenario 2.2 |  |  |  |  |  | X |
| Scenario 2.3 |  |  |  |  |  |  |
| Scenario 2.4 |  |  |  |  |  | X |
| New services, integrating into infrastructure | Scenario 3.1 |  |  |  |  |  |  |
| Scenario 3.2 |  |  |  |  |  |  |
| Scenario 3.3 |  |  |  |  |  |  |

Table 6: A key element of Scrum sprints are retrospections of the past sprint, and which (social) issues the team should work on to improve.

|  |  |  |
| --- | --- | --- |
| **Sprint retrospections** |  |  |
|  |  |  |
| **Sprint 1** |  |  |
| **Appriases** | **Criticisms** | **What to improve** |
| Keep review and planning in the same meeting | Clarify what needs to be documented where | Keep review and planning in the same meeting |
|  | The menaing of task states is unclear | Better documentation of the Sprint tools |
|  |  | Remove the "Delivered" state |
|  |  | Improve communication in sprint execution |
|  |  |  |
| **Sprint 2** |  |  |
| **Appriases** | **Criticisms** | **What to improve** |
| Scenarios are starting to give us more scope, CH institutes are committing. | Compile a list of actions in the minutes, the backlog seem not enough. Perhaps assign tasks to specific people in such way. | Compile a list of actions in the minutes, the backlog seem not enough. Perhaps assign tasks to specific people in such way. |
| Conference calls get more efficient using less time |  | Add conf call access details in the invitation Email |
|  |  | People should join timely to the conference call |
|  |  | Shorten conference call timeslots to 1.5 hours from now on |
|  |  |  |
| **Sprint 3** |  |  |
| **Appriases** | **Criticisms** | **What to improve** |
| Contributions from Eva and Rosette are good | Roberto has never shown up | Participation of WP members |
| Those participating were timely | Generally low participation in the WP5 conf calls | Not enough collaboration between WP3 and WP5 (Borje) |
|  | Borje as Product owner has never shown up |  |
|  |  |  |
|  |  |  |
| **Sprint 4** |  |  |
| **Appraises** | **Criticisms** | **What to improve** |
| Eva and Rosette are committed and proceeed on PoC1 | Much work was accomplished that was not captured in tasks | When finishing tasks, team members are allowed to adjust the points! |
|  | More fine-grained steps in defining the tasks | Everybody is encouraged to work with and edit the backlog |
|  |  | Once ytou change the backlog, drop a mail to WP5 notifying of this change |
|  |  |  |
|  |  |  |
| **Sprint 5** |  |  |
| **Appraises** | **Criticisms** | **What to improve** |
| Eva and Rosette are committed and proceeed on PoC1 |  | Better circulate the meeting minutes |
| Lajos joined today, and plans to do so in the future. |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **Sprint 6** |  |  |
| **Appraises** | **Criticisms** | **What to improve** |
| Eva and her e-Infrastructure reports are well received and appraised | Still low participation in the sprint planning meetings | Changing scenarios in practice in WP5 should be better synchronised and documented towards WP3 |
| Eva likes that she has fnished something! :-) |  |  |
| Meeting minutes are now circulated appropriately now. |  |  |
|  |  |  |
|  |  |  |
| **Sprint 7** |  |  |
| **Appraises** | **Criticisms** | **What to improve** |
| Eva and her e-Infrastructure reports are well received and appraised | We need to know when people go on vacation | Maintain people's vacation plans in the Backlog |
| Eva likes that she has fnished something! :-) |  | Michel to advertise untaken scenarios on DCH-RP mailing list |
|  |  |  |
|  |  |  |
|  |  |  |
| **Sprint 8** |  |  |
| **Appraises** | **Criticisms** | **What to improve** |
|  | Very slow responses from INFN and Roberto |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **Sprint 9** |  |  |
| **Appraises** | **Criticisms** | **What to improve** |
|  | lack of priorities for tasks | Be more "pushy" towards Rosette and Borje to provide priorisation |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **Sprint 11** |  |  |
| **Appraises** | **Criticisms** | **What to improve** |
| The SCRUM methodology was appreciated. | Borje was concerned in the beginning whether SCRUM would work. But during the actual PoC1 phase he was convinced that SCRUM is very successful. | Add a PoC report about using SCRUM to D5.3 |
|  | Scrum is about the management of a project, not the content. Scrum was not designed for this, so there is a need for a solution that facilitates the open communication between team members during the sprints. |  |
|  |  |  |
|  |  |  |

1. <https://documents.egi.eu/document/1544> [↑](#footnote-ref-1)
2. [https://docs.google.com/a/egi.eu/spreadsheet/ccc?key=0AuKxnKM\_liK-dGpqcGV6ZkVScDhGRjFzZ3gybWp0eVE - gid=0](https://docs.google.com/a/egi.eu/spreadsheet/ccc?key=0AuKxnKM_liK-dGpqcGV6ZkVScDhGRjFzZ3gybWp0eVE#gid=0) [↑](#footnote-ref-2)