# logo DCH-RP low resolution

# First Proof of Concept: Extended scenarios stemming from D3.1 and subsequent testing of components by WP 5

Document status: Draft

Document version: Version 0.10

Type (Internal or Restricted or Public):Internal

Document date: July 16 2013

Work packages: WP 3 supported by WP 4

Document reference: <https://documents.egi.eu/document/1677>

# Document change control

The table below includes the revision number, the date of update/issue, the author responsible for the changes, and a brief description of the context and/or scope of the changes in that revision.

|  |  |  |  |
| --- | --- | --- | --- |
| Revision Number | Date of Issue | Author(s) | Brief Description of Change |
| 0.1 | 06 March 2013 | R. Ruusalepp | Extended structure of the report |
| 0.2 | 07 March 2013 | B Justrell | Additional text |
| 0.3 | 08 March 2013 | B Justrell | Editing and adding text |
| 0.4 | 09 March 2013 | Börje Justrell | Editing and adding text |
| 0.5 | 11 March 2013 | M Dobreva | Additional text |
| 0.6 | 12 March 2013 | B Justrell | Additional text |
| 0.7 | 4 April 2013 | M Drescher | Editorials |
| 0.8 | 28 May 2013 | B Justrell | Additional scenario |
| 0.9 | 5 June 2013 | B. Justrell  R. Ruusalepp | Editing additional scenario |
| 0.10 | 16 July 2013 | C. Prandoni | Editorials |

## 1. The aim of the first proof of content

In the first proof of content, the aim is

* to identify representative situations that mirror the cultural heritage institutions archiving workflow problems in the field of digital preservation.
* to provode guidance on testing to what extent the e-Infrastructure (NREN, GRID and other cloud providers) can help to resolve these problems.
* To illustrate typical situations related to preservation which users (be it intermediaries – archivists, curators, or end users – researchers, citizens) would encounter.

In this document the idification of digital preservation situations has the form of scenarios on different level of concreteness. For the the actual proof of content a number of components has been identified to be piloted by WP5. Each component is devided into four distinct topics:

* What needs to be tested
* Possible procedures
* Desired outcomes
* Expected feed back to WP 3

## 2. Scenarios

Scenarios illustrating organisational challenges are identyfied and described in deliverable in D3.1 (reproduced here as Scenarios 1.1 -5 below). In this document these scenarios are expanded with

* scenarios illustrating end user concerns (e.g. researchers in digital humanities, curators, general citizens) and
* scenarios related to testing existing services and defining new ones (have the gaps described in D3.1 but could also refer to work done in KEEP and SCAPE projects).

### 2.1 Scenarios illustrating organisational challanges

##### Scenario 1.1. Using specialised research tools from a digital humanities e-Infrastructure on material preserved in-house

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.

##### 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.

##### 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.

##### 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.

##### 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.

### 2.2 Scenarios illustrating end user concerns

##### 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).

##### 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.

##### 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.

##### 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).

### Scenarios related to testing existing services and defining new ones

Scenarios related to testing existing services and defining new ones (have the gaps described in D3.1 but could also refer to work done in KEEP and SCAPE projects)

##### 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.

##### Scenario 3.1.a: New scenario focusing on trust

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 using for validating their objectives and methods as well as their management of intrinsic threats 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 criteria for the services it will get from the GRID.

##### 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.

##### 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.

## 3. Components to be piloted in WP5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Component** | **What needs to be tested** | **Possible procedures** | **Desired outcomes Expected feed back to WP3** | |
| **Ingest** | A cultural heritage institution wants to ingest a number of different record types to its grid-based preservation system. All files need to be checked for integrity and consistency with standards using automated routines that document the outcomes of checks. All files need fixity information to be attached to them, including persistent identifiers that would allow for identification and to check file integrity at any point in time. | Assemble a test sample of different file formats (ca 10 different record types with different file formats). Choose software tools for the required processes (integrity checking, file format checking, fixity information, persistent identifiers) that meet the requirements of archives.  Design installation / customisation process for the tools to run on the Grid platform.  Run tests.  Measure success of outcomes.  Measure time and effort required for the exercise. | Tools run without failures.  Processes run fast.  The integrity of all files can  be checked after the ingest  process.  The level of automation of  the entire process is high.  Time and effort required is  manageable. | To what extent tools for the  required ingest processes are In  place and how well they are  running. What are the  conditions for this (time and  effort required) |
|  |  |  |  |  |
| **Storage** | The grid-based preservation system has to store the files in a way that they can be retained with full accessibility and usability. The authenticity of the files should also be guarantied. Strategies for replacing obsolete technology with new technology have to be in place. | Use the test samples mentioned above for checking if raw data are fulfilling the requirements on formats and standards.  Check if appropriate metadata standards are in place.  Evaluate the strategy for inserting new technology.  Run tests  Measure success and outcome  Measure time and effort required for the exercise. | Requirements on formats  and standards for raw data  are fulfilled.  Appropriate metadata  standards are in place as  well as a trustworthy  strategy for replacing  obsolete technology.  Time and effort required is  manageable. | To what extent the require-  ments on storage are met.  What are the conditions for this  (time and effort required). |
|  |  |  |  |  |
|  |  |  |  |  |
| **Active digital preservation** | The grid-based preservation system has to have a number of complementary curation services like  - Schedule-based integrity checking  - Dereferencing and delete  - Migration and MoveOut of preserved files to new version of software and/or handware  - Possiblities to export data  - Conversion and transformation of data  - Administering retention | For the test samples, the data quality should be checked on bit-level to see if  - preservation requirements are met for different curation services  - the preservation outcome quality is acceptable according to standards or other requirements set up by the cultural heritage institutions  - integrity checks are done or possible to do (manually or automatically ) based on a fixed schedule.  The level of transparency (including documentation) in the administration of the retention activities has to be in accordance with the requirements of the cultural heritage institutions.  Run tests  Measure success and outcome  Measure time and effort required for the exercise. | Tools run without failures.  Curation services run fast  and meet the requirements.  Level of transparency is  acceptable.  The level of automation of  the entire process is high.  Time and effort required is  manageable. | How to evaluate to what extent an cultural heritage institution is mature to implement active digital preservation and what additional capacity it needs to develop in cases there are any defficiencies. |
|  |  |  |  |  |
| **Access** | Needed services are  - List items  - Find items  - Retrieve items  - Emulate  - Administration of access | The listed services should be checked on the test samples.  Each service should be acceptably met according to critical metrics for access services (response, time, precision, recall) set up by the cultural heritage institutions.  Run tests  Measure success and outcome  Measure time and effort required for the excercise. | Tools run without failur. To what extent services  for access are in place, and are running well .  Time and effort required  is manageable.  Matrix of metrics and minimum requirements for quality. | How cultural heritage institutions can select services meeting its needs, and how to select from available offers. |
|  |  |  |  |  |
| **Organisational issues** | There has to be clear agreements on outsourcing in place covering aspects like  Cost reduction (see separate row on Economic implications)  Increased effectiveness  Increased quality  Acceptable level of resources (technical and human)  Minimising risks  Policies in place | The service providers should present drafts on agreements that covers the services provided. | Draft text of agreement that both culrural heritage institutions and service providers has judged to be right or commendable. | How cultural heritage institutions can handle outsourcing situations. |
|  |  |  |  |  |
| **Services architecture** | Agreements on standars has to be in place that covers services lika  Data resource setup Interoperability,  Aggregation  Advanced search support  Persistent identifiers  User Authentication and access control | The service providers should present drafts on agreements on service architecture | Draft text of agreement that both culrural heritage institutions and service providers has judged to be right or commendable. | How to evaluate to what extent an cultural heritage institution is mature to evaluate what service architecture is needed. |
|  |  |  |  |  |