WP5: Task 56 (Describe testing procedures for scenario 1.3)

Authors:

Eva Toller (National Archives of Sweden, RA)
......

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
0.1	20130411	Eva Toller	RA	First draft
0.2	20130423	Eva Toller	RA	Added more potential (non-Swedish) cloud archive providers to investigate. Added an activity: test and evaluation of the ARC Graphical Client.
0.3	20130502	Eva Toller	RA	Added ROND as a tool to be tested (for de-identifying data).
0.4	20130520	Eva Toller	RA	Added three more tools for local testing.
0.5	20130807	Eva Toller	RA	Moved Xena from Scenario 2.2 (DCH-RP_WP5_Scen-2-2_ID-70.pdf) to this document and Scenario 1.3

1   TESTING PROCEDURES FOR SCENARIO 1.3

1.1   SCENARIO DESCRIPTION

"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."

General comment: if this scenario is reused in Proof of Concept #2, we could try to find a *real* organisation that has this problem (although it does not have to be a museum).

Suggested test data: see document **DCH-RP_WP5_Scen-1-3_ID-51.pdf**

1.2   SUGGESTED TEST PROCEDURES

1. Tools

   a. Investigate the existing tools that may be applicable for this scenario. Use the previous results from DC-NET a basis: http://digital-scholarship.org/dcrg/dcrg.htm and the subsequent results from DCH-RP WP 3.1: **D3.1_DCH_RP_ver_1.0_070213.doc**

   b. When feasible, test the chosen tools locally (at RA) and determine which ones that are useable/useful and if they should be included in the SweGrid/SweStore trials.

2. SweGrid/SweStore

   a. Investigate all preparations that must be done to make usage of SweGrid and Swestore

   http://snicdocs.nsc.liu.se/wiki/Getting_started_with_SweGrid

   http://snicdocs.nsc.liu.se/wiki/Swestore

   http://snicdocs.nsc.liu.se/wiki/Accessing_SweStore_national_storage_with_the_ARC_client

   b. Arrange a meeting with SweGrid/SweStore and determine what they can do for DCH-RP.

   c. Make all the necessary preparations and agreements with SweGrid/SweStore.

   d. Upload data to SweStore and test the chosen tools there.

   e. Test and evaluate the ARC Graphical Client.

3. Obtain permissions from RA to use and export data

   a. Obtain permission from the Electronic Archives section ("ElArk", with Christina Olsson as their representative) at RA to use the data outside the Preservation Net ("Bevarandenätet"), although still within RAs internal net – or, if needs must, on a computer not connected to the Internet.

   b. Select a number of appropriate DjVu images.

   c. Obtain permission from the Audio-Visual section ("L&B", with Magdalena Salomonsson as their representative) at RA to use the data outside the Preservation Net ("Bevarandenätet") and outside its customary storage place, although still within RAs internal net – or, if needs must, on a computer not connected to the Internet.

   d. Obtain permission from the legal owner of the information ("informationsägare", with Karin Åström-Iko as their representative) to use the data outside RA – that is, to use them in the SweGrid/SweStore infrastructures.

4. Cloud archive providers apart from SweStore (theoretical exercise only)

   This task has been moved to Proof of Concept #2 (PoC 2).

   a. Make an inventory of the cloud archive providers (CAP) in Sweden and other suitable alternatives (for example, http://publicdata.eu/ and http://ckan.org/)

   b. Compile a list of requirements appropriate for the scenario. Observe that the time horizon is 5 years according to the scenario. Examples of requirements: guarantee authenticity, migrate files to new formats when the old formats become obsolete.

   c. Construct a questionnaire and send it to the CAPs.

   d. Evaluate the answers and determine if a commercial CAP is suitable for a *small* institution like that in the scenario.

5. Miscellaneous

   a. Metadata: check the status of the Swedish eARD project for the chosen file formats.

1.3 DEPENDENCIES

The following chronological dependencies should be noted:

- 1b is dependent on 1a to be at least partly finished.

- 2b is dependent on 1b and 2a to be at least partly finished.

- 2c is dependent on 2b to be finished.

- 2d is dependent on 2c to be finished.

- 3c is dependent on 3b to be finished.

- 3d is dependent on 3a and 3c to be finished.

- 4c is dependent on 4a and 4b to be finished. **(Moved to PoC 2)**

- 4d is dependent on 4c to be finished. **(Moved to PoC 2)**

[A graphical description of the dependencies may be inserted here]

1.3.1   Activities that can start immediately

- 1a (Investigate the existing tools that may be applicable for this scenario…)

- 2a (Investigate all preparations that must be done to make usage of SweGrid and Swestore…)

- 3a (Obtain permission from the Electronic Archives section…)

- 3b (Select a number of appropriate DjVu images.)

- 4a (Make an inventory of the cloud archive providers (CAP) in Sweden.) **(Moved to PoC 2)**

- 4b (Compile a list of requirements appropriate for the scenario…) **(Moved to PoC 2)**

- 5a (Metadata: check the status of the Swedish eARD project for the chosen file formats.)

- 2e (Test and evaluate the ARC Graphical Client)

2 TOOLS TO BE TESTED

The choice of tools to be tested and evaluated will be guided by the following appeal from WP3:

"The developed solutions need to be tested for their simplicity of installation, management and use."

A fourth criterium will be used: generality of solution.

See **DCH-RP_WP5_Scen-1-3_ID-ToolTests.pdf** for the results of the tool tests. There, these four criterias will be graded on a scale from 1 (very bad) to 5 (very good).

For each tool, there is a (postulated) reason for why that tool has been chosen for tests.

2.1 ROND

ROND (Riksarkivet Open Data) is a tool for de-identifying data sets. It is dependent on Riksarkivet's chosen meta data format for structured text files, **ADDML** (http://xml.ra.se/addml/, in Swedish). ROND will be used on "Filmregistret".

Reason: to be able to use a collection of records in this scenario. Not needed for preservation, only for later dissemination. Thus, only tested for "meta-purposes" in this scenario, but the results should be relevant to other scenarios.

2.2 ARCHIVIST'S TOOLKIT

For download and information, see: http://archiviststoolkit.org/download/release/2_0

"The Archivists' Toolkit™, or the AT, is the first open source archival data management system to provide broad, integrated support for the management of archives. It is intended for a wide range of archival repositories. The main goals of the AT are to support archival processing and production of access instruments, promote data standardization, increase processing efficiency, and lower training costs."

Postulated reason: a simple, easy-to-use tool is needed for institutions like the one in scenario 1.3, with little or no IT competence.

2.3 DSPACE

For download and information, see: http://www.dspace.org/introducing

"DSpace is the software of choice for academic, non-profit, and commercial organizations building open digital repositories. It is free and easy to install "out of the box" and completely customizable to fit the needs of any organization. DSpace preserves and enables easy and open access to all types of digital content including text, images, moving images, mpegs and data sets. And with an ever-growing community of developers, committed to continuously expanding and improving the software, each DSpace installation benefits from the next."

Postulated reason: open source, easy to install, can be used for most types of contents.

2.4   FEDORA

For download and information, see: http://www.fedora-commons.org/about/features, http://sourceforge.net/projects/fez/

"The Flexible Extensible Digital Object Repository Architecture is a conceptual framework that uses a set of abstractions about digital information to provide the basis for software systems that can manage digital information. It provides the basis for ensuring long-term durability of the information, while making it directly available to be used in a variety of ways. It is very important to understand that Fedora provides a foundation upon which to build a variety of information management schemes for different use cases, not a full solution for a specific use case. The Fedora software that DuraSpace distributes has been designed to provide many different possibilities for a large array of applications."

Postulated reason: widely known (and well established?), but may be relatively hard to learn and use.

2.5   XENA

For download and information, see: http://xena.sourceforge.net/, http://sourceforge.net/projects/fez/

"Xena software aids digital preservation by performing two important tasks: detecting the file formats of digital objects converting digital objects into open formats for preservation."

Postulated reason: part of Digital Preservation Software Platform (http://dpsp.sourceforge.net/). Would be interesting to test but may not support all file formats that we need (especially DjVu).

Note: this tool was previously intended for Scenario 2.2. However, since it obviously can handle many more file types than images (and can only convert images to a very limited number of formats) it was obviously more suited for scenario 1.3.

< More tools may be added here if necessary >