

WP5: Task 70 (Describe testing procedures for scenario 2.2)

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	20130520	Eva Toller	RA	Added four tools for local testing.
0.3	20130613	Eva Toller	RA	Added a section on target formats.
0.4	20130624	Eva Toller	RA	Added a section on meta data.



1 TESTING PROCEDURES FOR SCENARIO 2.2

1.1 SCENARIO DESCRIPTION

“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.”

General comment: for test purposes, it may suffice to use 10 – 20 images. If this scenario is reused in Proof of Concept #2, a more realistic amount of images may be used, if possible.

Suggested test data: see document **DCH-RP_WP5_Scen-2-2_ID-66.doc**

1.2 SUGGESTED TEST PROCEDURES

1. Migration tools
 - a. Investigate the existing migration tools that may be applicable for this scenario (especially batch methods). Use the previous relevant results from DC-NET as a basis: <http://digital-scholarship.org/dcrg/dcrg.htm> and relevant 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.
3. Obtain permissions from RA to use and export data
 - a. Select a number of appropriate DjVu images.

- b. 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 their customary storage place, although still within RAs internal net – or, if needs must, on a computer not connected to the Internet.
 - c. Obtain permission from the legal owner of the information (“informationsägare”, with Karin Åström-Iko as their representative) to use the images outside RA – that is, to use them in the SweGrid/SweStore infrastructures.
4. New format(s)
- a. Decide what new file format(s) the images should be migrated to (in practice, this will be dependent on what migration tools that are available). If possible, the new format(s) should be present in the regulations of Riksarkivet, RA-FS 2009:2, see <http://www.riksarkivet.se/Sve/RAFS/Filer/RA-FS%202009-2.pdf>. If there are no tools that can migrate to acceptable formats, this should be noted in the evaluation.
 - b. Metadata: check the status of the Swedish eARD project for the chosen new file format(s).

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.
- 3b is dependent on 3a to be finished.
- 3c is dependent on 3b to be finished.
- 4a is dependent on 1a to be at least partly finished.
- 4b is dependent on 4a to be finished.

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

1.3.1 Activities that can start immediately

- 1a (Investigate the existing migration tools that may be applicable for this scenario...)
- 2a (Investigate all preparations that must be done to make usage of SweGrid and Swestore...)
- 3a (Select a number of appropriate DjVu images.)

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-2-2_ID-ToolTests.doc* 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 AVS DOCUMENT CONVERTER 2.2

For download and information, see: <http://www.avs4you.com/AVS-Document-Converter.aspx>

"Transfer regular text formats to e-Pub format and create your own e-books easily. Open and convert such e-book formats as DjVu and FB2 to all key formats supported by [AVS Document Converter](#)."

Postulated reason: supports conversion to/from DjVu (and many other formats).

2.2 UNIVERSAL DOCUMENT CONVERTER

For download and information, see: <http://www.print-driver.com/>

"Universal Document Converter is the most complete solution for the conversion of documents into PDF, JPEG, TIFF or other graphical files. The underlying basis of the program is the technology of virtual printing. As a result, exporting documents into a chosen format is not any more complicated than printing on a desktop printer."

Postulated reason: supports conversion to/from DjVu (and many other formats), said to be easy to use.

2.3 A-PDF DjVu to PDF

For download and information, see: "<http://www.a-pdf.com/djvu-to-pdf/>"

"A-PDF DjVu to PDF is a fast, affordable utility to allow you to batch convert DjVu (.djvu, déjà vu) into professional-quality documents in the PDF file format."

Postulated reason: batch conversion, supports conversion to/from DjVu (and many other formats). Not free software, but there is a free trial version.

2.4 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."



DCH-RP: Digital Cultural Heritage Roadmap for Preservation - Open Science Infrastructure for DCH in 2020

EC Grant agreement no: 312274

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).

< More tools may be added here if necessary >

3 TARGET FORMATS

As stated in section 1.2, the new format(s) should be present in the regulations of Riksarkivet, RA-FS 2009:2 (<http://www.riksarkivet.se/Sve/RAFS/Filer/RA-FS%202009-2.pdf>). The following formats are accepted:

- **JPEG** (Joint Photographic Experts Group), <http://www.jpeg.org/> , SS-ISO 10918
- **TIFF/IT** (Tag image file format for image technology),
<http://www.digitalpreservation.gov/formats/fdd/fdd000072.shtml> , ISO 12639:2004
- **PNG** (Portable Network Graphics), <http://www.w3.org/TR/2003/REC-PNG-20031110/> , ISO/IEC 15948:2004
- **PDF/A-1** (PDF 1.4, Portable Document Format), <http://www.pdfa.org/> , SS-ISO 19005-1:200

The fourth format (PDF/A-1) is only permitted for scanned images (thus, not for Born Digital ditto); however, since scenario 2.2 is all about *digitized* art images, PDF/A-1 is suitable.

It should be documented if it is impossible for *all* of the four chosen tools (listed in section 2) to make conversions to one or several of the four target formats. This may be an indication of the (un)suitability of a format for *de facto* usage, regardless of official regulations.

Scenario 2.2 is concerned with active digital preservation and usage (access), but it is not clear if the usage is about images that should be as detailed as possible (high resolution) or images that should use little bandwidth and have a “good enough” resolution. Thus, at least one of the target formats should be a high resolution format, and at least one (other) target format should be of “good enough” resolution only.

4 META DATA STANDARDS

Even if the amount of test data in this scenario is relatively small (about 200 images), there are many collections of images and other documents that easily grow to be huge, and can be counted in several thousands (sometimes up to several millions).

Swedish Riksarkivet (<http://www.riksarkivet.se/>) handles such big collections but has not yet decided on a final meta data standard to describe such collections.

A Swedish project, “e-Archiving and e-Records Management” (**eARD**) is concerned with meta data standards for various types of archives. However, the eARD project has not yet dealt with large collections of documents (including images). It is hoped that this will be done later, but it may also be decided that the Norwegian-Swedish meta data standard **ADDML** (<http://xml.ra.se/addml/> , in Swedish) will be used for document collections.