

WP5: Task 66 (Decide on test data for scenario 2.2)

Authors:

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

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

Revision History

Revision	Date	Author	Organisation	Description
0.1	20130409	Eva Toller	RA	First draft
0.2	20130430	Eva Toller	RA	More image types should be tested: jpeg and TIFF . Permission from %jud & Bild+only necessary for the jpeg images.
0.3	20130527	Eva Toller	RA	Documentation of the selected test data. This version of the document is restricted until permission is obtained to show the images in section 1.5 in public.

1 TEST DATA 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.

1.2 SUGGESTED TEST DATA

RA has a lot of digitised images in the **DjVu** format (<http://djvu.org/>). Although these images may not be as old as 15 years, they should be appropriate to use as test data for this scenario. Furthermore, conversion from the DjVu format is a real-life problem.

The chosen images should not have any security restrictions, neither should they have any copyright restrictions. The subject of the images should, if possible, be art images (or as close to that as possible).

2013-04-30: It has been proposed that some other types of images should also be tested: jpeg (<http://www.jpeg.org/>) and TIFF (Tagged Image File Format).

1.3 PREPARATORY WORK

General:

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

For the DjVu, jpeg and TIFF images:

- Select a number of appropriate DjVu images.
- For the jpeg images: Obtain permission from the Audio-Visual section (led by Magdalena Salomonsson as their representative) at RA to use the images 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.
- 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.

1.4 DEPENDENCIES

The following chronological dependencies should be observed:

- Before any practical work (except for administrative preparations) can start, permission from the Audio-Visual section must be obtained.
- Before any data can be uploaded and processed in the SweGrid/SweStore infrastructures, permission from the legal owner of the information must be obtained.

1.5 DESCRIPTION OF SELECTED TEST DATA

1.5.1 Jpeg images from the 'Digital commission' (from 2004)

The 'Digital commission' was a government agency committee. The data set consists of 12 selected jpeg images. These images were chosen for diversity, since most of the test data images for Scenario 2.2 is of DjVu type.

The images are mostly alike to 'holiday pictures', taken by people when travelling on behalf of the committee.



Figure 1: Example of jpeg image from the 'Digital commission' (cropped)

1.5.2 TIFF images (miscellaneous)

This data set consists of 7 selected TIFF images. Like the jpeg images, these images were chosen for diversity. The *motif* of the images are very diverse: page from a Church population registration book, a post card, a group picture of unknown people... A couple of the images are very large (150 and 397 MegaByte, respectively).



Figure 2: Example of TIFF image (post card)

1.5.3 Old geometric maps (DjVu format)

This data set consists of 21 selected DjVu images. They are the result of a national land survey in the 17th century. The images can be directly viewed here:

<http://www.riksarkivet.se/default.aspx?id=21561&refid=22519> ,

<http://www.riksarkivet.se/default.aspx?id=21573&refid=22519> , Östergötland, Memming & Hammarkinds h:d. %h:d+is short for %härad+, an old Swedish word for county.

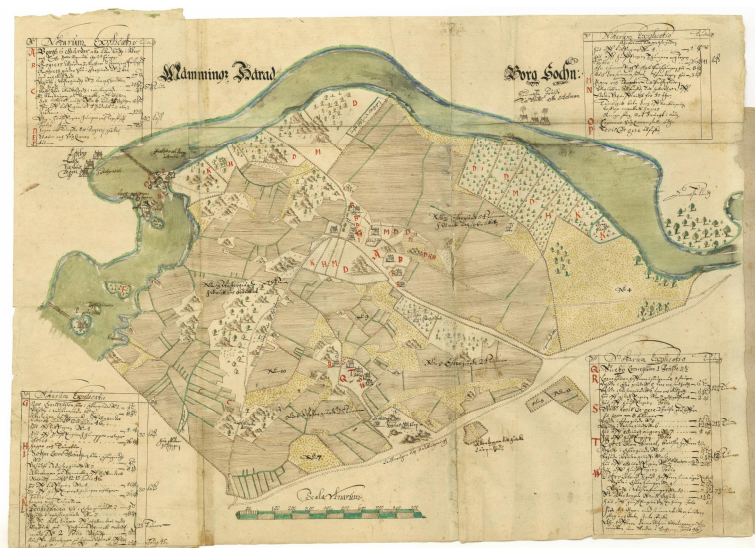


Figure 3: Example of DjVu map image: survey from Memmingshärad (county), Borgs socken (parish)

1.5.4 Magnus Erikssons landslag (DjVu format)

This data set consists of 19 selected DjVu images from Magnus Eriksson landslag (the rural law of Magnus Eriksson). Magnus Eriksson was a Swedish king reigning between 1319 and 1364.

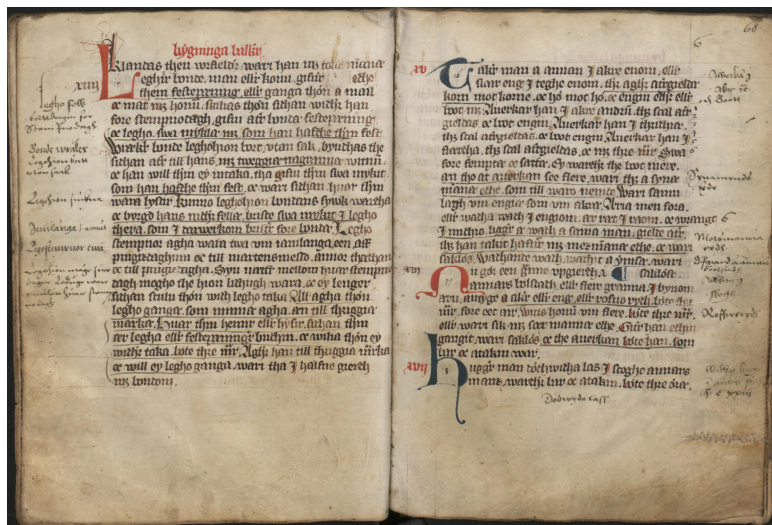


Figure 4: Example of DjVu image: Magnus Erikssons landslag

1.5.5 Pergament- och brevpapperssamlingen (the Parchment Collection, DjVu format)

This data set consists of 3 images from Letters of Indulgence (in Swedish, *avlat*) and 10 images of very old sheet music.



Figure 5: Example of DjVu image: Sheet music, *Graduale* (cropped)

1.5.6 Sketches and watercolours by Fredrik Arosenius and Johan Georg Arsenius (DjVu format)

This data set consists of 31 images by Fredrik Arosenius and 4 images of Johan Georg Arsenius, from the Military Archives at Riksarkivet. Both Arosenius and Arsenius were officers in the Swedish National Defence Force. They lived and worked during the 19th century.



Figure 6: Example of DjVu image: Sketch by Fredrik Arosenius

1.5.7 Maps from the Military Archives (DjVu format)

This data set consists of 14 maps from the Military Archives at Riksarkivet. They are all from the county of Östergötland.



Figure 7: Example of DjVu image: Map of the peninsula Ångölandet+