WP5: Task 68 (Decide on test data for scenario 2.4)

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	20130409	Eva Toller	RA	First draft
0.2	20130527	Eva Toller	RA	Documentation of the selected test data.

1 TEST DATA FOR SCENARIO 2.4

1.1 SCENARIO DESCRIPTION

"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)."

1.2 SUGGESTED TEST DATA

The test data should, naturally, be the archived web site itself. It consists of a number of html pages and other types of documents. However, the tests may use other types of storage models than the current one (and, if necessary, other web sites to test different tools for downloading) . The Linné site has no security restrictions.

1.3 PREPARATORY WORK

General:

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

For "Linnéjubiléet":

- Obtain permission from the Electronic Archives section ("ElArk", with Christina Olsson as their representative) at RA to use the web site outside the Preservation Net ("Bevarandenätet"), 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 web site outside RA – that is, to use it 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 Electronic Archives 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 The web site "Linnéjubiléet" (Linné Anniversary)

This web site was constructed for the 300th anniversary of the birth of Carl von Linné. It contains information about specific anniversary activities, about Linné and his life, about gardens and exhibitions, and much more. The site consists of 4058 files altogether, and the total size is 469 MegaByte.

The orginal site, http://www.linne2007.se/ , is not reachable any more (or at least, only itermittently). The Google cache of site has the URL http://webcache.googleusercontent.com/search?q=cache:ZzAIn4DYkSsJ:www.linne2007.se/+&cd=1&hl=sv&ct=clnk&gl=se .

The site was previously downloaded to Riksarkivet for preservation by HTTrack (http://www.httrack.com/). If more tools for downloading of web sites are to be tested (as suggested in **DCH-RP_WP5_Scen-2-4_ID-72.doc**), it should be preferable to test them *both* on the Google cache and on a "live" web site (http://www.linne2007.se/ if it can be reached, otherwise a similar but otherwise arbitrary web site).The reason for this is that a situation where Riksarkivet should preserve a site that is only reachable as a Google cache will very easily happen in "real life".
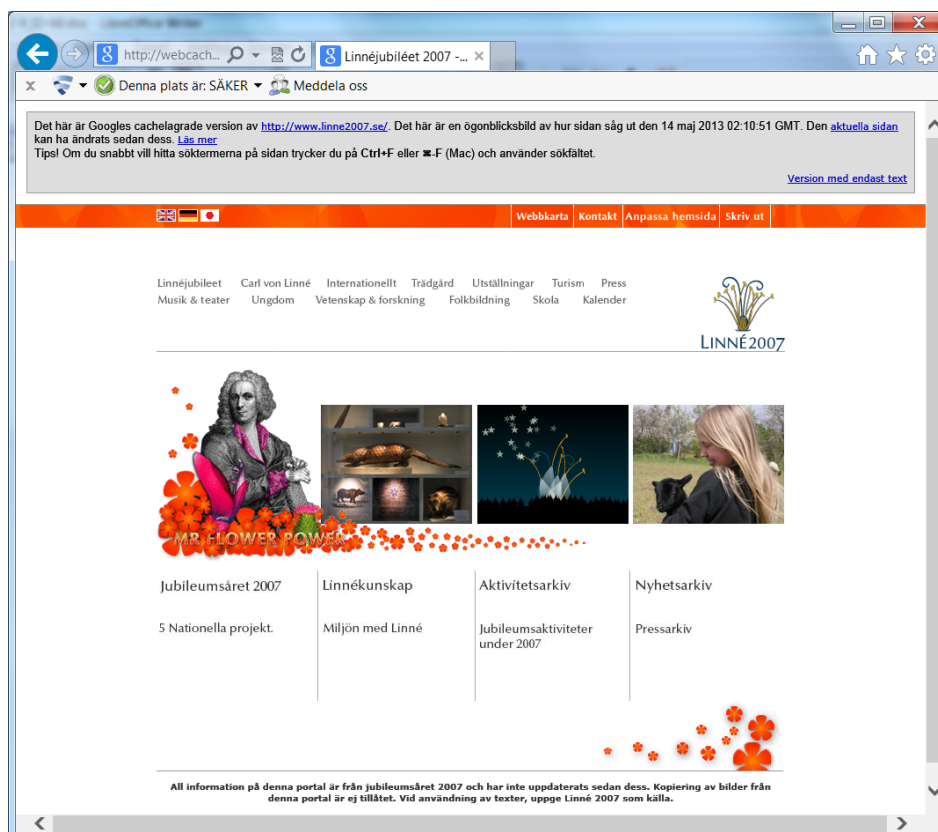

Figure 1: First side of Linné anniversary web site (from Google cache)