WP5: Digital Preservation at Cultural Heritage Institutions in Sweden:
Analysis of the Current Situation and of Future Needs and Requirements

Project Acronym: **DCH-RP**

Grant Agreement number: 312274

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

Revision: Draft 1.0

Authors:

Eva Toller (National Archives of Sweden, RA)
Sanja Halling (National Archives of Sweden/Digisam, RA)
Rolf Källman (National Archives of Sweden/Digisam, RA)

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	2014-05-04	Eva Toller	Riksarkivet	First draft
0.2	2014-05-14	Sanja Halling, Rolf Källman	Digisam/ Riksarkivet	Made corrections
0.3	2014-05-21	Eva Toller	Riksarkivet	Made more corrections
1.0	2014-06-16	Eva Toller	Riksarkivet	Made final changes

TABLE OF CONTENTS

1   EXECUTIVE SUMMARY

Digisam is a secretariat for National coordination of digitisation, digital preservation and digital access to cultural heritage, established by Swedish Government to coordinate the continued development work on digitisation issues and organised as a department of Riksarkivet (National Archives in Sweden, http://digisam.se/index.php/en). One of the main tasks of Digisam is connected to a proposal on how coordinated and cost-effective preservation of digital cultural heritage information at Swedish state cultural heritage institutions that collect, preserve, and provide access to cultural heritage material, should be designed.

As a first part of our work on this issue, Digisam has carried out a pilot study on long-term digital preservation. In this pilot study, conducted between April 2013 and March 2014, the current state of storage and preservation at cultural heritage authorities, was examined and compared with the current %state of the art+.

This document presents the results of the a pilot study. The results are based on in-depth interviews with seven cultural heritage government agencies, concerning their current processes and solutions for digital preservation and future needs concerning digital preservation.

2 INTRODUCTION

The new digital resources, in particular the ‰Web 2.0+and social media, have made it possible for cultural heritage institutions to make their collections available in a new and engaging way. However, in order to do so, the cultural heritage information must also be preserved in a reliable way.

The rapid development of technology decreases the sustainability of computer hardware and software to a few years. Although migration of stored cultural heritage information does not have to be made at the same rate, nevertheless this means that there is a need for continuous observation in order to ensure that measures can be taken in time to secure future access to cultural heritage information

2.1 OBJECTIVES OF THE DELIVERABLE

The objectives of this deliverable is to present the results of the interviews made with some Swedish cultural heritage institutions, the conclusions that can be made from the results, and what measures that should be taken in order to facilitate digital preservation in the cultural heritage sector in Sweden. The result may possibly be applicable to other countries than Sweden.

2.2 PARTICIPANTS IN THE PILOT STUDY

The interviewees were the following:

- Arkitektur- och designcentrum (The Centre for Architecture and Design, http://www.arkdes.se/, choose ‰Välj språk+for English version)

- Institutet för språk och folkminnen (Institute for Language and Folklore, http://www.sofi.se/2072/, short description in English)

- Tekniska museet (National Museum of Science and Technology, http://www.tekniskamuseet.se/1/start_en.html)

- Riksantikvarieämbetet(Swedish National Heritage Board, http://www.raa.se/om-riksantikvarieambetet/in-english/)

- Statens historiska museer (National Historical Museums, http://www.shmm.se/default____160.aspx)

- Statens maritima museer (National Maritime Museums, http://www.maritima.se/en/)

- Kungliga Biblioteket (National Library of Sweden, http://www.kb.se/hjalp/english/)

The answers from the interview with the National Library of Sweden have not been included in this context because the library sector needs to be analysed deeper, in collaboration with the National Library.

2.3 MAIN CONCLUSIONS

The three following issues were considered most important as initial measures:

- Co-operation concerning preservation issues;

- More general/common solutions for digital preservation;

- A centralised preservation solution for government cultural heritage institutions (including archives and libraries).

3   DIGITAL CULTURAL HERITAGE INFORMATION

The amount of digital born cultural heritage information (%born digital+) is rapidly increasing. Therefore, it is of vital importance that secure and reliable digital preservation of data and metadata plays a natural part at the cultural heritage institutions. Henceforth, the term %cultural heritage information+ (from now on referred to as *CHI*) is used (if not otherwise stated) to include both information about cultural heritage data (that is, metadata) and the cultural heritage data collections themselves. In order not to confuse this with cultural heritage *institutions*, the latter are always abbreviated as %CH institutions+. %CH+ always means %cultural heritage+.

3.1   STORAGE VERSUS PRESERVATION

In the context of this study, *storage* is defined as purely technical storage (on bit level). *Preservation* is defined as securing the stored information so that future users can reach it, read/see/hear it, and understand it.

In order to secure the CHI, it is important that the long-term preservation issues are taken into account already when the information is created. Consequently, at the point of digitisation of information, issues as (for example) format conversion and storage media should be taken into account. Otherwise, there is a risk that information will be destroyed and/or inaccessible due to (for example) faulty format conversion or storage media failure.

Within the scope of the pilot study, different levels of preservation were defined:

- Short-term preservation . solutions that are used for a short time, 5 years maximum.
- Medium-term preservation . solutions that are used during a system's lifetime, 10 years maximum.
- Long-term preservation . solutions that are used after the originating system's lifetime, number of years unspecified.

The interviewees often had large amount of digital CHI, but no systematic long-term preservation, and seldom specific systems for long-term preservation. There was a general concern about continually increased amounts of digital CHI, which will consequently bring on higher costs for both storage and for preservation.

3.2   METADATA AND RAW DATA

The collections at the six CH institutions are described in formats deriving from the traditional (analogue) way of describing collections, and international standards for metadata are seldom fully implemented. Therefore, metadata structures differ between the institutions.

As a contrast, archival and library information (both metadata and raw data) is described and preserved according to established principles and international standards, and there are systems for long-term preservation.

There is an overall ambition that all CHI (archival information, metadata about collections, and raw data) should be preserved for the long-term future. However, in practice there are differences in how archival information and metadata about collections are treated. For the most part, it is not clear if the latter should be (legally) treated as archival information or not (and how, and *if*, it should be preserved for long-term). Metadata about collections is often modified and migrated between different systems, mostly without the long-term preservation perspective. Therefore, there is a risk that information will be lost.

3.3 PRIORITIES, ROUTINES, AND STRATEGIES

Today, the CH institutions have no support for making priorities for digital preservation. Furthermore, there is a need for developing routines and strategies for long-term digital preservation. The first step is establishing plans for digitisation of the collections at the institutions; existing work is supported by Digisam. However, a more general support for work concerning digital storage and preservation is desired, as well as clarification of roles in digital preservation (which responsibilities the administrators and managers of information, and their counterparts for systems, should have).

Some Swedish CH institutions will participate in the DCH-RP test of EUDAT services (http://eudat.eu/) , in order to contribute to the validation of the roadmap, and also to examine if the roadmap can be adapted to a national version. The aim is to integrate digital preservation in all digitisation steps.

3.4 COMMON REQUIREMENTS

A majority of the interviewees are in favour of a centralised storage/preservation solution within the CH sector. A common example was a support function for storage, where (for example) file conversion tools (especially for *mass conversion*) can be accessed and/or downloaded. Another example of support functions is handbooks. Yet other examples are help with planning long-term preservation and information about international standards.

The majority is in favour of a central (government) storage solution within the CH sector where, for example, the National Archives/Digisam is the main participant. As part of the ongoing collaboration between the National Archives and the government agency *Statens servicecenter* (National Service Centre, http://www.statenssc.se) around cost-effective e-archive services, it may also be possible to include storage solutions for different kind of cultural heritage data. For example, there may be a common function for bit-level storage. If a common support function for long-term preservation of digital CHI is included, the services may include support for planning long-term preservation, accepted international standards and how these can be implemented, tools for mass conversion/migration, et c.

A first step towards improvement of long-term preservation may be to harmonise metadata standards. In Sweden, a recent initiative has been taken with the **FGS**'s for archives (**F**örvaltnings**g**emensamma **s**pecifikationer, which approximately translates to "Common Specifications for Government Agencies", see http://riksarkivet.se/Media/pdf-filer/Projekt/eARD_informationstext_eng.pdf). FGS's have been developed for case management, personnel records, and the overall package structure for e-archives. The next FGS's to be developed are for archival descriptions, databases and registers, and file collections. For the CH information sector, corresponding FGS's could be centrally developed and administrated.

The *OAIS* model (**O**pen **A**rchival **I**nformation **S**ystem) is often used within the archival and library sectors as a conceptual model, and may be suitable also for other types of CHI.

As a consequence of the analysis of the interview results, and of discussions with many CH institutions, Digisam has suggested the following approach:

- A proposal for a centrally managed and administrated support resource for the CH sector.

- Adaptation and implementation of DCH-RP's *Roadmap for Preservation* for Sweden.

- Assessment of the suitability of centralised and/or distributed storage/preservation solutions for the CH sector.

- Forms for cooperation on storage/preservation issues among the CH institutions.

4 RESULTS FROM QUESTIONNAIRE

A summary of the answers to the questionnaire is presented in this section.

Query #1a: What type of cultural heritage information does your institution have?

Most of the institutions have several types of digital CHI. All interviewed institutions have information about collections, about artefacts, and about archives; all of them also have digital images and audio-visual material. Almost half of the interviewed institutions also have 3D drawings/ blueprints. Some of the institutions are currently planning digitisation projects and/or collection projects that include 3D information. The increase of 3D data is partly problematic, since routines and processes to preserve this type of CHI are not yet developed and/or standardised.

Query #1b: How does your institution manage your cultural heritage information (in general)?

Most commonly, metadata is gathered into one or several separate collection information systems, while digital CHI is stored in folders on one or several in-house servers or personal computers. The interviewees see a need for a more centralised management, but only one institution has actual plans for moving their digital CHI to a cloud solution. Most of the institutions have less than 10 Terabyte of data.

The systems for collection information do not always support international standards for metadata, and this is regarded as a problem.

Query #2: Are there different priorities for what types of cultural heritage information that should be stored/preserved?

Four of the interviewees answer "no" to this question. However, in practise there is material that get a higher priority for digitisation: CHI that is regulated by law, and CHI that runs a high risk of being destroyed due to bad quality, old age, et c.

Those who answered "yes" added that they do not make a distinction between storage and preservation.

Query #3: Do you have a documented strategy for preservation of cultural heritage information (short-term, medium-term, long-term)?

One interviewee had a documented strategy for short-term and medium-term preservation. None of the interviewees had currently a documented strategy for long-term preservation.

Query #4: Do you have routines/procedures for digital storage/preservation of cultural heritage information?

Two of the institutions have implemented routines for digital storage and/or long-term digital preservation (examples: appraisal). However, some CHI previously digitised and stored on CDs, are not included in these routines and thus risk to be overlooked/forgotten.

Query #5: Do you have systems for digital storage/preservation of cultural heritage information?

All of the six institutions have systems for digital storage. However, it was a bit unclear what exactly comprises a %system+, but it was agreed that it should be some type of well-structured storage solution.

Only two of the institutions have systems for long-term digital preservation. In both cases, the collection information system also serves as a long-term preservation system. As for the rest of the interviewees, they expressed hopes that the manufacturer of the collection information system also would solve the problem with long-term preservation.

Query #6: Are your systems adequate for your current needs, with respect to digital storage/preservation of cultural heritage information?

Only one of the institutions answered that their current needs were covered, with the reservation that their future needs would *not* be.

Query #7: What types of metadata standards do you use to describe digital cultural heritage information that will be stored/preserved?

Except for within the archival and library sectors, the metadata standards that are used differs widely. Within the museum sector, standards are often developed in-house and/or are built on old traditions. International, established standards are used less often. However, when a new system for collection information is implemented, this is taken as an opportunity to change metadata standard.

Query #8: Do you use classification for any of the stored/preserved information (with respect to confidentiality, availability, traceability, consistency et c)?

All of the six institutions had information that needed classification. Some of this is processed at least partly automatically (for example, availability for different types of users). Sometimes the processing is altogether manual (for example, ethical considerations whether to publish or not).

Query #9: Are there different levels of authorisation for the digital cultural heritage information (especially for access)?

Five of the six institutions had different levels of authorisation for access/usage.

Query #10: Are there different requirements for the quality of the digital cultural heritage information, depending on the type of material?

One institution had different requirements for quality. This is mostly for digital objects. There seems to be no such different requirements for metadata.

Query #11: Are you using cloud services for storage and/or long-term preservation?

None of the interviewees are currently using cloud solutions for storage and/or long-term preservation of digital CHI. One institution aims to store high-resolution images in a central (internal) cloud, connected to the collection information system.

Query #12: What tools, services and automated processes are used today? What would be needed in the future?

Currently, at least five of the institutions use tools, services and/or automated processes for the following:

- Storage

- To supplement metadata

- To search CHI (both data and metadata)

- Administration of events

- Authorisation

There was no consensus regarding what tools, services and automated processes that would be mostly needed in the future. However, the following four was most often mentioned (written in order %most needed+first):

- Solutions for storage (preferably common/centralised solutions, or at least solutions shared with others)

- Standards (including the previously mentioned FGS's).

- Routines and automated processes for mass conversion/mass migration

- Guides and %toolkits+

Others that were mentioned were the following six:

- Extraction of information

- Classification of information

- Format identification

- Packaging

- Validation

- Appraisal

Query #13: Do you think that a common (shared) solution for storage/long-term preservation is needed for cultural heritage information?

Five of the interviewees answered that this is needed. One of the interviewees answered that before such a solution is implemented, official standards must be established.

The institutions were sceptical to cloud solutions, especially when the information is controlled by commercial parties, or is stored abroad. Instead, a solution controlled by the government is preferred.

Query #14: Do you think that a common (shared) support function is needed for cultural heritage information?

All interviewees answered that such a function is needed for storage and long-term preservation. The function should include general support, advice, et c. It was also pointed out that advices and recommendations are not sufficient: regulations are also needed.

Query #15: What other means for supporting storage/long-term preservation of cultural heritage information would be needed in future?

Strategies for conversion/migration was mentioned as an important support function.

4.1 CO-OPERATION WITHIN THE CULTURAL HERITAGE SECTOR

As seen in the answers to queries #13 and #14 (and partly also from answers to queries # 12 and #15), there is a great need for common solutions, including storage and (international) standards, and co-operation in general.

Some institutions/projects in Sweden and Europe that participate in co-operation within the CH sector are enumerated below.

Sweden:

- LDB-centrum (Centre for Long-term Digital Preservation, http://www.ltu.se/centres/Centrum-for-langsiktigt-digitalt-bevarande-LDB?l=en)

- eARD (the e-archive and e-diarium project, http://riksarkivet.se/Media/pdf-filer/Projekt/eARD_informationstext_eng.pdf)

Europe:

- Digital Preservation Europe (DPE, http://www.digitalpreservationeurope.eu)

- Open Planets Foundation (http://www.openplanetsfoundation.org)

- APARSEN (http://www.alliancepermanentaccess.org)

- ENUMERATE (http://www.enumerate.eu)

- DCH-RP (http://www.dch-rp.eu)

- EUDAT (http://www.eudat.eu)

5 CONCLUSIONS

The amount of information that should be stored/preserved is rapidly growing within the CH sector. The institutions are not well prepared for this accumulation of data (however, the archive institutions seem to be somewhat better prepared than the museums). In the European perspective, the Swedish institutions are also somewhat less prepared than their European counterparts (again, this especially pertains to the museums).

A conclusion is that a roadmap for preservation (like the one that DCH-RP is developing) is of utmost importance for the Swedish CH sector: for defining what institutions should include in the preservation work, for deciding what measures that should be taken for a successful long-term preservation, and for how to proceed in general. Furthermore, common, shared, and preferably centralised resources are needed, both for cost-efficiency and for a higher quality of the information that is stored and preserved.

When common resources are implemented, they can be used to test the roadmap for preservation. However, it should not be forgotten that preservation is not the final objective; the real success of the roadmap is fulfilled when citizens, companies, government agencies, scientists and researchers, schools, and developers can easily access and use the digital CHI.