Denmark

Jakob H. Petersen
The Danish Library Agency

The Danish digitisation efforts in regard to cultural heritage are coordinated by the Danish Ministry of Culture. In response to the communication from the Commission in the fall of 2006 the ministry has appointed a committee to work on a national strategy for digitisation. The work is focusing on digitisation for improved access, digitisation for preservation, and copyright and ipr-issues in regard to digitisation. The present progress report is prepared by the the Danish Library Agency in collaboration with the major institutions in the relevant areas.

A. Users and content

Digitisation in Denmark started in 1990 with registries and catalogues thus making metadata on collections machine readable. The more common use of Internet from the mid-nineties meant that electronic registries and catalogues were a way to make the existence of the collections visible to the general public.
Today there are six major national portals which utilise these digitised metadata in addition to other data:

- <http://www.library.dk>, The Danish Union Catalogue with metadata on materials in Danish libraries
- <http://www.kulturarv.dk/mussam>, The National Database on artefacts and other materials in Danish Museums
- <http://www.kid.dk>, The National Database on Art in Danish Museums
- <http://www.arkivalieronline.dk/>, Church registers and Population Censuses on the Internet. The service offers full text data
- <http://www.dkconline.dk>, The National Database on archaeological sites and monuments in Denmark
- <http://www.kulturarv.dk/fbb>, The National Registry of Listed Buildings.

All six services are extensively used by the general public. In addition to these national services there are a wide range of other internet-based services that provide access to digitised information.

Text
Since most registries and catalogues are already digitised the focus in digitisation is on the materials themselves rather than on metadata. According to an analysis from 2005 text is by far the largest category of material being digitised (Projekt Digitalisering og filmning – kortlægning og analyseresultater, Biblioteksstyrelsen maj 2005, <http://www.bs.dk/showfile.aspx?IdGuid={41C34E10-B976-4285-8970-1EE9252C5FA2>.
There has been extensive digitisation of sound primarily at The State and University Library but it is expected to decrease relative to digitisation of moving images.
The institutions responsible for the largest share of text digitisation are The Royal Library and The Danish State Archives Filming Centre. The digitisation of text at The Royal Library has primarily included journals, books and manuscripts. The Danish State Archives Filming Centre has digitised church registers and population censuses. Digitisation of text is almost exclusively done by scanning but a small part of the materials have been digitised by means of manual transcription/inputting. A small percentage of the scanned text has been ocr-processed. There has been a large number of specific text digitisation projects at other institutions – the largest being the digitisation at The State and University Library of manuscripts from Danish radio news.
In addition to common text digitisation some text materials – primarily newspapers and archival materials – are scanned to microfilm. The primary purpose of this activity is preservation. The State and University Library and The Danish State Archives Filming Centre are the most active institutions in this area.

The Royal Library, E-Resources
<http://www.kb.dk/en/materialer/e-ressourcer/index.html?subject=>

Archive for Danish Literature
<http://adl.dk/adl_pub/forside/cv/forside.xsql?nnoc=adl_pub>

Sound
Digitisation of sound is done both for preservation purposes and in order to increase access to materials. The Danish National Library for the Blind, The State and University Library, and The Danish Broadcasting Corporation (DR) are doing most of the digitisation of sound. The digitisation of music and subsequent conversion to mp3 files by The State and University Library has formed the basis for online distribution of music by both libraries and commercial entities.

The State and University Library, music downloads
<https://<http://www.bibliotekernesnetmusik.dk/netmusik2006/>

Images
Images include a wide variety of categories. In addition to photographs, paintings, and drawings the digitised images include books, maps, handwritings, and sheet music. The Danish National Gallery digitises art in order to improve access to the materials and facilitate research. The digitisation is also done for purposes of registration and in order to represent the materials in print. Digitisation of images for print is also done at The Royal Library but the main purpose of digitisation of images at the library is to improve access to the materials or protect the materials.

The Royal Library, National Library Pictures
<http://www.kb.dk/en/nb/materialer/billeder/index.html>

Moving images
The digitisation of moving images has hitherto been a limited activity. Most of the work is being carried out by The Danish Broadcasting Corporation (DR) and The Danish Film Institute. The digitisation is primarily done by external partners and the main purpose of the digitisation is to improve access to the materials. The State and University Library has done some digitisation of video film (older television and cinema advertisements). The Danish Broadcasting Corporation is continuously digitising materials on demand from its own archives for use in the DR production.

The State and University Library, television and cinema advertisements
<http://www.statsbiblioteket.dk/danskreklamefilm/eng/forside_eng.php>

Objects
Digitisation of objects is primarily done by museums. There is a distinction in this area between primary digitisation (of the objects themselves) and secondary digitisation (digitisation of metadata such as registries and catalogues). As there are still catalogues and metadata in this area that haven't been digitised the secondary digitisation continues to be the most comprehensive activity. The National Cultural Heritage Agency is conducting most of the digitisation on behalf of the museums. A notable result so far is the availability to the public of information on more than 1,3 mil. artefacts via "The Museums' Collections" on <http://www.kulturarv.dk/mussam>. Similarly the National Cultural Heritage Agency is responsible for <http://www.kid.dk> which offers online access to 39.000 works of art in Danish museums and collections. The National Museum has built a database containing data on more than 750.000 objects and a database with digital pictures of 200.000 of these objects. The Danish National Gallery has a similar database.
There is some work being done with 3D light scanning of sculptures at The National Museum and The Danish National Gallery but the activities are at a very early stage.
In regard to sites and monuments protection the digitisation has so far primarily been done to increase efficiency of registration and administration. The national sites and monuments database (<http://www.dkconline.dk>), which is run by the Danish National Cultural Heritage, presently holds information on 170.000 Danish localities Recently, the Danish National Cultural Heritage Agency has established a database of protected buildings with digitised photographs (see. <http://www.kulturarv.dk/fbb>).

Distribution of content
Apart from the national inventories mentioned above the main method of distribution is subject or institution specific mediation on the Internet. The relevant common denominator for digitised content is rarely the fact that it has been through a digitisation process but the materials' relationship to a specific collection, institution or theme. It is this relationship that forms the basis of the mediation to the end user.
This means that digitised content is often used in relation to a subject based or thematic presentation. A good example is the presentation of the digitised version of "The Danish Music Journal" from the subject based internet gateway "Online Music Research Library" rather than its presentation along with other digitised journals. Some of the digitised materials are presented through very specific internet gateways as this allows searching and presentation to be tailor made for the digitised material. A more comprehensive digitisation of primary materials should probably be supplemented by national gateways to the information or by the inclusion of the materials in existing national gateways.

E-learning or creative industry uses for digitised material
The increased emphasis on student centred independent learning outside the classroom and the use of learning materials not specified in the curriculum means that mass digitised content is of general relevance to modern education.
The Danish Broadcasting Corporation provides the most notable example of e-learning use of digitised content. Their platform <http://www.dr.dk/skole> offers a subscription based access for all Danish primary schools to multimedia content for education. The platform has been extended to include high school level education and there are plans for the inclusion of university level education as well. The initiative is supported by the Danish Ministry of Education but the use of a subscription model for supplementary revenue is an interesting financial model for digitisation. The subscription model also has interesting technical and distributional aspects as it is based on cooperation in those areas and the possibility of digitisation on demand.
A similar if not as advanced example is <http://www.e-museum.dk> where schools can find information on how to integrate the collections of Danish museums in the education.
Another example is the delivery of selected advertising films by the State and University Library to UNI-C's service for primary schools <http://reklame.vhostip5.isop.net.uni-c.dk/> based on a special agreement with COPY-DAN.
The commercial use of digitised materials is primarily inspirational. There is, however, one example of a very specific commercial use of digitised material. As part of the Ministry of Cultures effort to support legal use of music it was decided to provide a better alternative to illegal download of music. The State and university library's digitisation of Danish music was used as basis for the establishment of a service that offers both library loans of MP3-files and purchase through commercial services <http://www.bibliotekernesnetmusik.dk>.

The general accessibility of the digitised cultural content
Apart from the obvious barrier presented by copyright legislation the most important issue in regard to accessibility is to ensure that the end user is aware of the material. This means that materials must be visible from national inventories or through more general commercial search engines.

B. Technologies for digitisation

Standards and Technology by categories of material	Technologies	Standards
- objects		
- text documents	Image scanning, OCR and XML mark up	TIFF 6.0 400 dpi Grey tone
- images	Image scanning	TIFF 6.0/JPEG 400 dpi Grey tone
- audio	Digitisation (Dalet 5.1)	Preservation format: BWF, wav Browser format: MP3
- moving images*	Digitisation (ARDCAP)	Preservation standards: MPEG2 or MPEG1, DVCPRO50 (DR)

*The Danish Film Institute (DFI) remarks that the information is primarily concerned with Broadcast/Standard Definition. DFI considers Digital Betacam as the best preservation format for Standard Definition. In regard to movies there is no consensus on preservation format but HDSR is expected to be the chosen format for movies and jpeg2000 for HD and digital movies.

Interoperability
The most important aspects of interoperability are the adoption of open standards and protocols that support meta-data frameworks and facilitate information exchange; for instance:

- Dublin Core based metadata that can by accessed from other systems or presented independently of the digitised materials.
- IT-architectures supporting SOAP-based web-services
- XML based sector-specific standards for object- og collections level metadata
- WAI-compliance

A working group consisting of experts from the MLA authorities and institutions in Denmark has developed a common format based on the Dublin core. The format, expressed as an XML schema, can be found at <http://www.kulturarv.dk/tjenester/publikationer/emneopdelt/kulturarv-it/abm2006/index.jsp>.
The general strategy for interoperability is metadata exchange through OAI-PMH and service exchange through the use of webservices.
In the library sector initiatives are ongoing to create a common access architecture based on integrated search, i.e. syndication of data from heterogeneous data sources.
DR uses the DR Metadata standard. However the standard specifies an interface for business to business exchange based on the Dublin Core Metadata schema. The formal specification for this interface is provided in the form of xml w3c schema. Concerning the audio and video files a new multi transcoding application is on the agenda. <http://www.dr.dk/0mDR/Metadata/20061211095304>.

Research needs
Digitisation of especially the audiovisual cultural heritage is of utmost importance, not just because the av-media are vulnerable and thus in danger of disappearing, but because access to the rich amount of sound and moving images is an indispensable source for research and study of the cultural and political history of the 20th century.
For all categories of material research is required to support an automatic structured metadata extraction. Work on the output of an OCR process is promising, but work on e.g. the markup of output from newspapers is still lacking. The area of automatic feature extraction from audio- and video material is still in the beginning stages and much research is still needed.
The area of how to present the results of a potential mass digitalisation process also needs further investigation. Finally, the whole area of digital preservation needs continued investigation.

C. Sustainability of content

Funding models and sources
Digitisation efforts in Denmark has been characterised by a number of separate projects rather than a continuous activity in mass digitisation.
The digitisation projects have typically been funded by a grant from the Ministry of Culture for either

a specific project or a specific institution.
The responsibility for digitisation projects has typically rested with the institution that owns the collection.
The potential drawbacks of this approach are the absence of an overall prioritisation between collections or materials and the risk of suboptimal use of digitisation machinery or outsourcing. On the other hand this project model ensures co-funding from institutions, a prioritisation that reflects local needs and thus greater sustainability.

Examples of cost-reduction
There are of course huge savings on internal processes when registries and catalogues are digitised. A good example is digitisation of the libraries' card catalogues that allows for much more efficient internal handling of materials.
There is also a benefit for the end user in terms of efficient searching, presentation and ordering of the materials in the catalogues.
The internal costs and the costs for the end user are to some extent hidden costs which can make it difficult to document cost-reduction.
The digitisation of the materials themselves entails a reduction in handling costs in relation to the usage of the materials by end users. As there are few examples of mass digitisation of entire categories of materials it is difficult to assess the full potential in cost-reduction for digitisation.
The libraries' migration from printed scientific journals to electronic subscriptions could probably provide examples of reduced handling costs and reduced costs in terms of time consumption for end users.
For instance the number of electronic loans from The Royal Library rose from app. 550.000 in 2001 to app. 1.900.000 in 2005. The library's staff was downsized considerably in this period and a increase in traditional loans of this size would not have been possible.

National competence centres in digitisation of cultural content
There are no official Danish competence centres in digitisation but the highlighting of activities by specific institutions in different material categories under (A) provide an indication of specific competencies.

D. Digital preservation

National policies for digital preservation
The general policy of the Royal Library and the Stae and University library is to enter into international consortia to address the universal problem of digital preservation. As a result, the two libraries are presently involved in two EU project on digital preservation and is active in IIPC. Denmark's act on legal deposit also includes digital materials. In addition to submission of materials from publishers the Danish part of the Internet is harvested and preserved. The Royal Library and The State and University Library are responsiblefor these tasks.
Also Danish radio and television programmes have become subject to legal deposit. Since July 2005 the State and University Library has recorded broadcast materials "off-air" The recorded programmes are automatically converted to a digital format and stored on the library's server for long time preservation. Further information is available on Netarchive.dk <http://netarkivet.dk/index-en.php>.

Examples of tools and technologies for digital preservation of Cultural Heritage content
Active bit preservation: In connection with the Net archive-software an application has been developed which supports bit-preservation of the harvested material. The software will be turned into open source. Integrity Check in connection with ingest: As part of a Danish (DEFF) activity around Fedora, the State and University library has developed tools to perform integrity check of the incoming material. The software will be available as open source software. Digital object management systems (DOMS): Both the Royal Library and the State and University Library are active in the process of creating DOMS, which support to long term preservation of digital material. The focus at the Royal Library has been on the output of their digitalisation process. The focus at the State and University Library has been a more generic model due to the greater diversity of material and based on Fedora.

E. Monitoring progress

The information on the quantitative aspects of digitisation in regard to input and output reflects the total digitisation effort over a longer time period and not just activities in 2006. The numbers can be viewed as

an approximation primarily based on information from a ten year period (1996-2006). There were digitisation efforts before 1996, but the bulk of the activities are post 1996. The approximation is based exclusively on numbers from the larger Danish institutions which means that they probably underestimate both input and output. It will hopefully be possible to deliver a revised and more complete overview in the next progress report.

Input/Output Indicators

Input	**MEUR / Posts**	**Comment**
Funding allocated to support digitisation - government	4 MEUR	Estimate 1996-2006
Funding allocated from within existing institutional budgets*	5 MEUR	Estimate 1996-2006
Size of work-force (full-time equivalent posts)	N/A	Could be estimated by subtracting technology investments from total cost and dividing with average wage. Many tasks were, however, outsourced.

*Combined total, including personnel costs.

Output	**Number of collections / items**	**Comment**
Number of digitised collections		Information is provided on total number of items rather than collections. The number of collections can be estimated by diving with the average size of a collection
- museums	22.500 pages, 228.000 photos, 15.000 maps	Numbers from The National Museum and The Danish National Gallery
- libraries	2,3 mil. Pages, 532.000 hrs. of sound and 320 hrs. of film and 100.000 photos (The Royal Library)	Numbers from The Royal Library and The State and University Library
- archives	6,3 mil. Pages 2.100 drawings,	Materials at The Danish State Archives - primarily censuses and registries
- Other	3000 hrs. of film and 4.800 hrs. of audio tape	The Danish Broadcasting Corporation and The Danish Film Institute
Number of digitised collections recorded in a national inventory		
- museums	2.1 mil. records	Object records at The National Museum and The Danish National Gallery and other recognized Danish museum
- libraries	2,5 mil. Records	Retrodigitisation of book catalogues at research libraries funded by DEFF 1998-2006
- archives	6,3 mil. Records	Church registers and Population Censuses
Number of digitised items per collection type		
- object entries (m/i)	2.1 mio object metadata	Object records at The National Museum and The Danish National Gallery and other recognized Danish museum

Output	Number of collections / items	Comment
- text documents (m/i)	8,7 mil. Pages, 8,8 mil. Metadata	
- images (m/i)	228.000 images with metadata	
- audio (m/i)	532.000 hrs. of sound with metadata	The State and University Library
- moving images (m/i)	3300 hrs. of film with metadata	The State and University Library, The Danish Film Institute and The Danish Broadcasting Corporation

Use indicators

Use	Amount	Comment
Use of web-sites distributing digitised heritage content		
- user sessions	4 mill. pr year.	Estimated use of the digitised material. Including usage from the national services (under A) and similar services that include other materials as well as digitised content would increase this number
- average duration of user sessions	N/A	
- average page impressions per user session	N/A	
Purchase or distribution of CDs / DVDs	None	
Audiences for streaming, video-casting or broadcasting	The Danish Broadcasting Corporation delivers streaming media to all Danish primary schools and most high schools. In addition some digitised content is offered to the general public	

Assessment of user-need surveys on digital heritage content services

There have been no specific surveys on the need for more specific kinds of digitised content but a lot of work has been done in determining user needs in general. The best documentation of user needs is the subsequent use of digitised materials and this seems to indicate a strong demand in many areas as especially younger users expects materials to be available online.
A study of user preferences based on the everyday life of the end user: <http://www.statsbiblioteket.dk/publ/fieldstudies.pdf>.