

## **Digitization of copyright protected newspapers in Denmark**

**Tonny Skovgård Jensen - State and University Library Aarhus**

In Denmark we have legal deposit of newspapers and have had it since the 18<sup>th</sup> century. We are two national libraries who each have a copy of the printed newspapers. Statsbiblioteket / The State and University Library in Aarhus have the main responsibility for the national collection of newspapers. We have around 23.000 meters of shelves with newspapers. The daily newspapers are estimated to cover around 74 million pages. Furthermore we have additional local newspapers mostly distributed weekly. Around 54 million pages of daily newspapers have been microfilmed since the 1950'ies. These microfilms are being copied for access at our own library as well as at The Royal Library in Copenhagen, and at other libraries buying copies. We are currently running a very large scale digitization project with the aim of digitizing 32 million pages from microfilms. This project is scheduled to be finished in the beginning of 2017.

### **The usage of the digitized newspapers**

After digitizing the printed copy at The Royal Library in Copenhagen is going to be discarded, that just one printed copy of each newspaper remains. The funding of the digitization is closely linked to this. The new digitized copy is used as a new and safe copy, which is ensuring the long term integrity of the collection together with the others copies. Consequently the first and most important use of the digitized pages is for the purpose of long term preservation. By January 2016 we have digitized around 17 million pages. About 10 million pages are available in our online newspaper portal Mediastream: <http://www2.statsbiblioteket.dk/mediastream/avis>. This is of course another important use: To provide better access to the content of the national newspapers collection.

### **The legal framework and copyright regulations**

By law we are allowed to provide everyone access to legal deposit newspaper content, but only when users are visiting our library or The Royal Library in Copenhagen. The period of copyright protection is defined by the lifetime of the individual journalists. For practical reasons we are using the publication date to define the year of online access by everyone: 100 years old or older newspapers are online available for everybody without any restrictions. In Denmark we have extended the collective licensing legislation. However it has not been possible to make an agreement for all the newspapers still being published.

### **Enabling research e. g. via TDM facilities**

We have just opened a new facility that we call the cultural heritage cluster. This is a high performance computer, which is strong enough to do big data analysis on the large quantity of data that we have in our digital collections. By having it on site our legal deposit library is able to offer services, which wouldn't be available otherwise. Indeed there are legal considerations still. Every project poses new challenges and questions regarding the legal framework for datamining.

### **The access for a wider public and the national responsibility**

For the full collection of digital newspapers the only way to provide access for every citizen is to provide access at the library premises. We expect that there will be about 2.5 million pages that are freely online available for everyone, because they are old enough to be outside of copyright protection. It hasn't been possible to negotiate extended collective licensing agreements for all of the copyright protected content. The publishers were not ready to make an agreement at this point, since several publishers are planning to make their own archives accessible. Supporting these activities we have offered a copy of the newspapers that we digitize for the publishers. Then the publishers themselves can make the copies available for the public. Several publishers have used this opportunity. By supporting the publishers in this way we are able to improve the access options of possible users. For newspapers that are copyright protected still, but out of print without any owner, we are working on an extended collective licensing agreement. This will make up

to 8 million pages available on site at the local public libraries countrywide and online remote access for research.

**Conflicting interests (economical, privacy, public, rights holder etc.)**

We are allowed by law to digitize our collections of printed newspapers for preservation. As well we are allowed to provide public access on site to the digitized pages too. The extended collective licensing option in Denmark makes it possible to make agreements for various kinds of access to copyright protected content. However these agreements cost money with the result, that the access is more an economical than a legal issue. The agreements are not mandatory for the right holders. We are aware of the risk, that right holders refuse an agreement. This is in fact what has happened with the newspapers that are still in print.

**What is “a good enough quality”? Which formats?**

We have chosen two angles on this: use and digital preservation. The digital images have to be usable, but not necessarily perfect. The quality of the digital image is good enough, if you can read the text at 100% display understanding the message. This also means that OCR quality is far from perfect. The OCR we have is still a great help for searching, even if it's not perfect. These choices reflect the kind of balance that most mass digitization projects need to find, between the quantity and the quality, given the fact that there are always financial restraints. The second angle is long term preservation. We have analyzed this quite in depth. Like many other preserving institutions we have chosen JPEG 2000 as our preservation format and a range of xml-files for the metadata.

**The tasks of the libraries**

Our main task is to keep the collections safe and accessible for a long time perspective. Digitisation is a key activity of this work. We see our obligation to promote this and to enlarge further dissemination of our collections, wherever it's possible. Working with the publishers and the right holders is one approach to implement this. Ensuring infrastructures, procedures, legislation for datamining and online access is the other approach. Many further ways to use the digital content has to be explored and will be unfolded.