

## **Guidelines for digitizing a newspaper**

Newspapers are loaded with information, but finding the information can be a challenge. Unless the newspaper has been microfilmed (and often even if it has been), libraries hold very few copies and these are often not available through interlibrary loan. Even with access to a newspaper, unless the paper has been indexed, searching for the desired information can be a tiresome quest.

Because of their vast amounts of hard-to-access information, local newspapers are often suggested as prime candidates for digitization. At its best, a digitized newspaper makes all its information searchable and accessible to anyone with internet access.

But newspapers present some distinct challenges on the road to a successful outcome.

The following document is intended to help illuminate many of the issues involved in newspaper digitization with the hope that

### **Why digitize?**

Document digitization can potentially greatly increase the breadth of access that can be provided. Digitized documents can be made available to multiple local users, and if desired, anyone with internet access.

Use of a digital surrogate may decrease the need to use the original – which may be fragile or difficult to access.

Digitized text documents have the potential of being keyword searchable.

### **What to digitize?**

What is not yet digitized – Not many newspapers have been digitized yet, but before embarking on a project it is necessary to make sure nobody has or is currently digitizing this same title.

What is legal to digitize – If the newspaper you want to digitize falls under copyright protection, you will need to create an agreement with the legal owner.

What is complete (or reasonably so) – Due to the nature of newspapers, anything less than reasonably complete makes for unsatisfactory use.

What is good quality – Digitizing cannot make a bad original look good. Digital images should be taken from the best available original.

What will be used – Digitizing newspapers is not an inexpensive endeavor and so you should make every effort to be sure that there is sufficient demand for the final product.

### **Digital and microfilm and print**

Microfilm remains the accepted standard for preserving newspapers. If the newspaper to be digitized has not yet been filmed, then filming should be the first priority. Filming and digitizing can happen simultaneously.

Newspaper digitization is most commonly done from microfilm, but there are reasons to use print.

Microfilm quality is a huge challenge – especially with older film. Digitization cannot compensate for film that is blurry or where the image is incomplete.

Best to use copy of earliest generation master one can find. (The film the was in the camera is the first generation master negative.)

Some large newspaper digitization projects are refilming old papers because the original film is not good enough.

Scanning from microfilm is generally faster and less expensive.

Some newspaper digitization projects have found better success image quality and optical character recognition success scanning partially or entirely from the original print copy.

### **OCR and metadata**

Once the digital image has been created they need to be made searchable and accessible.

The images are processed with optical character recognition software to make the newspaper text searchable.

Metadata must be created for each title, issue, image, and possibly article to help the users find what they are looking for.

### **Access and preservation**

Need to determine how you will provide ongoing access to this new resource. Some vendors also offer web hosting services.

CD/DVDs are not considered acceptable for long-term data preservation.

### **Vendor or in-house**

Newspaper digitization is a challenging process that is most often contracted out to a vendor.

Some libraries have done in-house newspaper digitization projects.

Treat this vendor as you would any other vendor that is looking for your business – they may say that they can do the work, but can they show you examples of successful projects?

Most vendors can create metadata to your specifications.

The Library of Michigan maintains a list of digitization vendors, many who have worked on Michigan projects. (This list may not include every vendor, and is not an endorsement.)

[http://www.michigan.gov/mde/0,4615,7-140-54504\\_50206\\_18643-146554--,00.html](http://www.michigan.gov/mde/0,4615,7-140-54504_50206_18643-146554--,00.html)

### **National Digital Newspaper Program**

The National Digital Newspaper Program (NDNP) run by the Library of Congress and the National Endowment for the Humanities is funding statewide newspaper digitization projects and has developed newspaper digitization guidelines

<http://www.loc.gov/ndnp/>

The content created by NDNP partners is housed at Chronicling America

<http://chroniclingamerica.loc.gov/>