4 Big Challenges in Downloading Historical Newspaper Articles

I have been researching online historical newspaper sites for several years, both the free ones as well as the subscription-based sites. Their software capabilities, independent of the size and quality of their scanned collections, break down into two parts:

1. How do you search the site and find "stuff"
2. How do you download and save the articles that you find

This post is all about the second part – how do you download and save materials.

If you remember only ONE thing about this Research Tip it is this:

"Make sure you plan ahead before you download an article."

Many of these sites use different underlying base software and they all are quite different in their approach to providing a download capability.

Among the varying features that I have found will require you to do one or a combination of the items below:

- Download a .pdf of the page that contains your desired article

- Download a .jpg of the page that contains your desired article

- Use a software snipping tool such as the Windows "Snipping Tool" or some commercial offerings, such as SnagIt, Shutter, Jing, etc. There are many of these tools. So you just snip the article with this tool.

- The newspaper library software will snip the article for you and present it as a "whole" article.

- The newspaper library software will snip the article for you and present it in several pieces that comprise the entire article, such as the headline, and each paragraph as an independent .jpg file

The concern with all of these methods is

"How do you end up with a complete article that is large enough to read or is zoomable?"

"Zoomability" is one of the biggest complaints that I hear from my fellow online historical newspaper researchers after they download an article.

The 4 challenges regarding downloading an article and ending up with what you want and need are presented below

Highlights – almost every online site presents the selected article after a search with highlighted search terms. So, depending on the capabilities provided by the site software, make sure that if you do NOT want to download the article with the highlighted text, that you download the article the way that you want it. There is always a way to get the article without the highlights. Experiment with the software or just download the whole page and crop your article later.

Chopped Up Articles – A few of the sites will present the article in a "chopped up" format, depending on the length of the article. Usually the title or headline is separated from the text, and if it is a long article, there may be several parts. So just like your need to concern yourself with downloading highlighted text, you will need to be careful with downloading articles that are broken up. It is safest to download the article (or entire page) in .pdf format and crop to your liking later. With .pdf formats you can generally zoom to the size you desire and then crop or snip the article in one piece.

The Article is Too Small – this is the challenge that requires the most forethought when saving the article for later use. Much of the time, the repository software will allow you to download or snip articles in several different ways. Unfortunately, if you do not check your downloaded image before you leave the site, you might be disappointed in the size or the quality of said download. Make sure that it is either zoomable after the download or if not, that the article is zoomed to a readable size prior to downloading. Also, if it is too small when downloaded, depending on how the site prepares the clipping, zooming later may degrade the quality so that when zoomed, it is so fuzzy that you can't read it.

Just like the other challenges, forethought will lead to the best results. It is always safest to download the article (or entire page) in .pdf format and crop to your liking later.

Oops, I Forgot the Citation – This is easily the biggest mistake one can make and creates many "smack your head" moments. When I first started searching newspapers, I got so excited when I found an interesting article that I just downloaded the article and did not record the details about the newspaper that I found it in. In order for me to find the article again for those that I discovered online – I will have to perform all those searches again. Many of the repositories have a function or a link where the source details are presented. What I do is include the publication title, publication date and page number as part of the file name of the downloaded .pdf or image file.

The moral of the story is that online newspaper collections can be an incredibly valuable tool for your research. Fortunately, much of what is written above and the challenges presented are not huge to overcome. Just think ahead and make sure that the articles that you download are readable and in a format and size that pleases you and your future readers. And make sure you document where the articles came from!

The easiest way is to save the entire page as a pdf and include the citation info in the file name.