



Legal Information Preservation Alliance

Digitization Is Possible Webinar Q&A February 6, 2015

1. Can someone explain the connection between resolution and image quality?

Erik Beck: **Resolution** refers to the concentration of pixels in the digital object. It is found by dividing the length of the image's diagonal expressed in pixels by the same length in inches. In image that has a high pixel per inch resolution will appear clearer and more detailed. **Image quality** in the context of still-image digitization, refers to the readability of a text. The QI measurement of a digital text, or its quality index, is a comparison of the height of the text's characters to the resolution of the digital image. The QI for color and grayscale images is found by multiplying the resolution (ppi) by the character height (h) in inches and dividing by 2:

$$QI = (dpi \times h) / 2$$

By most standards, an acceptable QI is anything 8 or above.

2. How do you deal with newsprint that bleeds through the page? Is there a scanner or tool to deal with this?

EB: If print from the succeeding page is showing through, you may use a blank sheet of paper to back the page you are scanning and block the following page. If you are scanning a page with text on both sides and print from the back is showing on the front of the page, I recommend capturing in grayscale or color so that the color depth of the printing on the front of the page will be greater than the bleed-through from the back. You may also use image editing software that will recognize bleed-through and eliminate it automatically. This operation is called "despeckling."

3. Would you be able to provide us with a list of quality (or vetted) OCR software?

EB: [ABBYY Fine Reader](#) is the best I've encountered. Its OCR function is extremely accurate. It's also very good at reproducing fonts and formatting in a rich text manifestation. It offers you an interface to edit the text it generates, and you can work with a wide variety of formats. [Tesseract](#) I've heard is a very good open source option. I strongly recommend avoiding Adobe Acrobat's text recognition functions for anything but text that was printed from a word processor.

4. Can you recommend a good scanner for documents and images?

EB: We at the University of Colorado Law Library use the [ATIZ BookScanner Pro](#). This is probably the least expensive scanner option on the market that is still capable of generating archival-quality images. We purchase the cameras to mount on the scanning rig separately. There are other higher end options out there from [Zeutschel](#) and [Bookeye](#) that you could consider.

5. Do you also use Digital Commons for a scholarship repository or do you use Content DM for scholarship as well?

EB: The University of Colorado Law Library has a Digital Commons repository that we use to house all works generated by our law school. It acts as our institutional repository. We find the Digital Commons network to be a valuable resource for publicizing our faculty's scholarship. However, Digital Commons is very restrictive in how it allows you to document your assets and customize your system. For our collections of primary legal material we maintain a separate digital library with the [Digital Collections of Colorado](#) and build special search features that integrate with that repository.

6. Must scholarship and other digital assets such as images be on the same DAMS?

EB: See response to #5.

Valeri Craigle: You can host your digital assets in a number of different systems, but you want to make sure they are OAI-PMH (Open Access Initiatives – Protocol for Metadata Harvesting) compliant and allow you to use a metadata schema that conforms to vetted standards (Dublin Core, for instance). We host our historical photographs collections in CONTENTdm, our E-journal publications in Open Journal Systems (OJS), and our Utah Supreme Court Briefs collections in Bepress at BYU.

7. Is anyone using PTFS Archival Ware?

EB: No, but thanks for bringing it to our attention. I was not previously aware of that product.

VC: The Utah State Library is using it for their "Utah Government Digital Library": <http://publications.utah.gov/>. Ray Matthews would be the contact for them if you have questions about the system. He can be reached at raymatthews@utah.gov

8. How have you prioritized particular digitization projects? How do you decide what to digitize next?

VC: Two criteria that we've used in choosing what to digitize include digitizing collections that are relevant at the time – such as our historical photographs collections at the time of our Centennial, and digitizing collections that are in danger of being lost or destroyed because of their condition. Right now, we are focusing on trying to digitize one of our faculty's collections of papers, as he has just retired recently and is in poor health.