History on the
Web / L'Histoire
sur la toile

The I.T. Insights of Historians

In 2014, the CHA Bulletin included a number of articles discussing the career prospects for recent history PhD graduates. Robert Talbot reviewed the recent American statistical study by L. Maren Wood and Robert B. Townsend examining the career paths of graduates, and noted that the rise in non-tenure-track positions is creating increased instability within the academic job market. Brandon Dimmel shared his own observations and experiences as a sessional instructor, and described the imperative of securing additional work to supplement academic income. Still another article discussed the concerns of "extradepartmental historians."

I completed my doctorate at Western University in the spring of 2014 and, like my peers, I am seeking a tenure-track position. As a SSHRC postdoctoral fellow at Trent University, I present and publish regularly and seek teaching opportunities. My attempts to network and find additional income sources have also led me to take the somewhat unusual step of co-founding Waterloo Innovations - a small business dedicated to research technologies.

This decision spurred directly from my graduate studies. The advent of inexpensive digital cameras and flash drives contributed to something of a revolution in archival research. Instead of transcribing the contents of documents or paying for expensive photocopies, researchers can now sign a reproduction permission form and photograph entire pages with a single shutter snap. Research that required weeks to complete thirty years ago can now be accomplished in several days for a fraction of the accommodation costs.

Despite these advantages, researchers who use digital cameras or scanners inevitably discover a problem. They return home with hundreds, thousands, or even tens of thousands, of JPGs™ with ambiguous filenames like "IMG_6392," and are forced to spend hours or even days sorting them all into groups with more meaningful titles. Researchers who want to annotate their photographs or run them through optical character recognition (OCR) software also have to individually input each set of images into PDF™ programs such as Acrobat Pro. Those who put-off this daunting and mundane task inevitably loose key documents in the mountains of images that they accumulate.

After performing this mechanical work for several years, I began to search for a way to automate the process. First, I hired a programmer to write a script for Acrobat Pro that created multipage PDFs from a series of file folders. Within few months, several of my fellow graduate students wanted copies. Installing the script, however, took half an hour and required users to change Acrobat Pro's settings and program files. Even after the installation was complete, users still had to manually separate their reams of images into file folders before running the script.

Given the unsolicited interest in this advancement, as well as my own desire to save time during the latter stages of my doctoral research, I contacted a software engineer from Silicon Valley named Douglas Wiebe, and we began to develop a more capable and user-friendly program that could automate post-archival photograph sorting. After two years of developing the software via Skype on Saturday afternoons, we built a website and marketed the product as Confero.

Researchers can purchase Confero 2.0 in advance of their research trip at our website waterlooinnovations.com. It works on Mac and PC. After finishing with an archival file, users take a picture of a custom QR code that they either generate and display on their smartphone, or print from Confero. This image serves as a bookmark that Confero can distinguish from the other pictures. When our customers get to their hotel room, or return from their research trip, they don't have to manually sort their images. Instead, they run Confero and it uses these "bookmark" images to separate the files back into the archival groupings while users check their emails, write an article, or watch TV. When Confero finishes, each user previews all of the image sets and file names in a matter of minutes (see picture). They then click "export" and go back to work or depart for the evening while Confero creates multipage PDFs and file folders of each sets' JPGs. OCRing these PDFs, importing them into reference managers like Endnote or Zotero, or sharing the same files with a research team on a cloud service only requires a few additional clicks. Hours or days of drudgery are reduced to minutes of keyboard time.

While producing and marketing Confero, I've enjoyed many unexpected opportunities. During the past year, for example, several history departments and the Wilson Center in Washington DC, invited me to instruct graduate students about research photography, managing dataflows and tweaking reference managers to organize and annotate PDFs. I've always found the students to be attentive, and the tutorials often turn into seminars, with everyone sharing methodological tips.

While working as a postdoc at Trent's interdisciplinary Environmental Studies department this past year, I also learned that historians' technological struggles are not unique. Archaeologists, geographers and biologists, for example, also use cameras, critter cams or microscopes for their research. I know researchers who literally spent months manually sorting hundreds of thousands of images into scientific datasets. Douglas and I consequently expanded Confero's sorting criteria to use GPS coordinates or time intervals to do the same work in minutes. Researchers who need to perform more sophisticated sorts can export their images' metadata (EXIF data) to a CSV spreadsheet, input their own formulas, and then use Confero to produce PDFs and file folders that only contain the desired images. Despite the science requirements that spurred us to develop these additional features, historians are, ironically, using the same functions to automate some of the steps involved in scanning old photocopies of archival documents.

These experiences demonstrate the benefits of interdisciplinary cooperation. I am in awe of historians who write their own programs. I could not have created Confero with my rudimentary programming knowledge. Yet Douglas was not aware of the historical community's requirements and would not have created Confero without my input. Historians are often just as well situated as other academics to appreciate and capitalize on the technological problems faced by 21st century researchers.

Daniel Heidt (danielheidt@gmail.com, @daniel_heidt) is a SSHRC postdoctoral fellow at the Frost Centre for Canadian Studies and Indigenous Studies, Trent University.

Douglas Wiebe (douglaswiebe@gmail.com) is a software engineer at Moody's in San Francisco and co-founder of Waterloo Innovations.