Gabrielle Sudik
IT1025 CRN 12115 – Spring 2017
Professor Tamerlano

Class Project:

Using Technology in History Research – What Works, What Doesn't, and the Role of the Historian

May 12, 2017

Introduction

Historical research has long been a painstaking and time-consuming endeavor. By its nature, historical information is old, sometimes hard to read or translate, often incomplete, and usually housed in a library or archive far from the researcher's home. Even learning about the existence of primary historical documents can be challenging, as they are often unknown or uncatalogued, let alone someplace where they can be studied easily. Sometimes, much-wanted documents are just lost to the ravages of time.

But advances in technology in the past two decades have greatly widened the scope of historical research, making it more accessible and thorough, and open to topics previously difficult to research. Library and archive card catalogs are available on the World Wide Web. Millions of documents, previously housed in basements of universities and governments, are now online. The full text of thousands of secondary sources is widely available from any computer – sometimes for a small cost, but often for free, especially through any research institution like a college or government agency. More importantly, primary sources are now more readily available for viewing by those who need them, even from distant locations, thanks to scanning projects, photographs, and scientifically-supported restoration and preservation efforts. Translations are also more readily available, either through technological means or thanks to the work of volunteers. Finally, technology is making the writings of historians more thorough, more accessible to readers, and more comprehensive.

This paper will first discuss what historical research generally requires, and why that was such a burdensome task until the last 20 years or so. Next, the paper will discuss how technology makes the task of historical research both easier and creates a better product, for both the scholar and the audience. Finally, the paper will discuss some limitations of technology and why the role of the historian himself is still essential in analyzing and writing about history.

What Is Historical Research?

A historian is a person who researches, studies and writes about the past, and is regarded as an authority, usually on a certain geographical area, era or topic. Historians learn the basic facts and standard opinions on a particular historical subject by reading what other historians have already written on the subject (secondary sources). Then they research letters, contemporaneous writings, photographs, government publications, collections of data, oral

narratives, maps, and artifacts like art or archaeological finds (primary sources) to contribute their own theories, opinions and discoveries on the subject. These findings are then published in a variety of forms, such as books, articles, websites, reports, museum scripts and exhibits, and documentaries.

Secondary sources are works (books, articles, documentaries) written by historians on a topic, presenting facts, theories and opinions. All historical research starts with secondary sources, so a historian first learns about the topic of his research and understands what has already been written about the subject. A good historian will understand which secondary sources are considered fundamental, and which are not, either because they lack scholarly rigor or do not contribute anything new to the scholarship. The beginning stages of historical research – when a historian familiarizes himself with others have already written – has long been a fairly simple step in the process, even without technology. Not so long ago, paper card catalogs and proprietary electronic databases allowed researchers to find books and articles on relevant topics. Bibliographies in the secondary sources often alerted a historian to other secondary sources. However, most card catalogs and electronic databases were usually limited to materials on hand, and to one language, or perhaps to two or three other major European languages. So searches would not reveal many items published in other countries. Similarly, even book bibliographies rarely referenced works written in another language, because getting ahold of foreign publications *and* translating them was quite difficult.

Researching primary sources has always been more difficult. Primary sources are the actual objects of history: Letters; first-hand accounts of events published in newspapers; government records such as laws, census results, maps, birth and death certificates, etc.; contemporaneous writings like scientific research results; photographs and video; and even tangible objects such as art, archaeological finds, scientific equipment, clothes, household objects, and weapons, etc. Until very recently, there was no easy way to replicate these objects and share them widely. There were rare exceptions for very famous works – like the millions of color copies of the Declaration of Independence or the oft re-printed photos of Lee Harvey Oswald's assassination – but most historical objects were not of high enough general interest to replicate and reproduce in large numbers.

So a researcher who was interested in, say, the government correspondence of a small 17th Century French village would have to travel to wherever these documents were housed, usually would need permission to see them, and may not be able to replicate them but would

have to take notes. Even knowing about the sources would have been difficult, for a researcher would have to rely on some sort of archivist or librarian to catalog the items and make it known publicly that they exist. Finally, as with secondary sources, there would be the issue of translating from one language to another. And in many cases, the issue of transcribing old handwriting into modern, readable text.

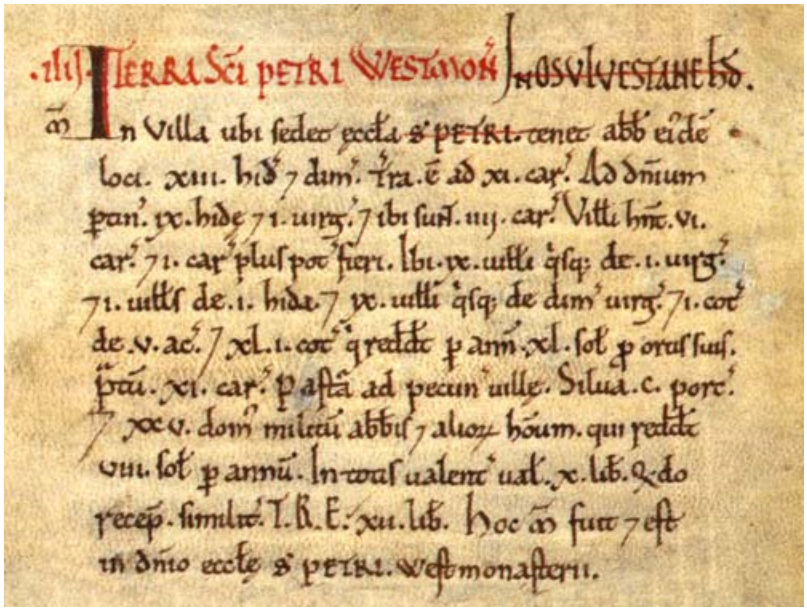
How Today's Technology Assists in Historical Research

Just a few decades ago, secondary source research meant physically visiting library after library in search of the best secondary sources.

With the advent of the internet and online availability of scholarly publications, research of secondary sources is now much simpler and more effective. No longer does a scholar have to travel to a distant library to read a rare journal – a copy of the needed article is quite likely online, or can be ordered and sent electronically, sometimes for a small fee. Online access to publications, including reviews of books if not the books themselves, allows scholars to more quickly weed out which publications are not needed, another feature of modern research that saves time.

Similarly, although finding works in foreign languages and, more importantly, *translating* works in other languages is still a challenge, a historian at least has access to those publications for a way that did not exist a few decades ago. And with the help of modern translating software, a historian can determine if a foreign-language article is even relevant. Even a rather poor translation will communicate the main topics and theories of an article, and the historian can then decide whether to pursue a proper translation or decide the piece is not relevant.

Primary source research has become even more accessible. Imagine, for example, an American historian who wished to write about the Domesday Book, a giant tome of English property and tax records written in 1086. There is only one copy of the book, in the British National Archives, in England. It is written in shorthand Latin, and in an archaic, non-standard font. Here is a sample page:



Until the past few years, a historian may have been lucky enough to find summaries of the Domesday Book (secondary sources), or possibly a very rare word-for-word translation of some sections. Even rarer might be photographs of some pages, but certainly not of the whole book. And it would have been terribly difficult to view the book itself, since it is (1) in England, (2) locked away in the National Archives and (3) not available for general perusal. Unless the researcher was lucky enough to find a translation from the Latin, he would have to both transcribe the ancient handwriting into modern letters, including the shorthand, and would then have to translate the Latin into English.

Today, however, the entire contents of the Domesday Book are available in a searchable, translated page on the internet.[1] Select photographs of the Book itself are also available online.[2] Although an entire scan of the book does not appear to be available, most any historian would find these two online resources sufficient, as together they cover both the book's entire content and examples of how it was written.

In other words, today's technology allows historians access to (at least some) primary sources in a number of ways: First, no travel is necessary. Second, many works have already been translated and transcribed. Third, documents that have been converted to searchable text are, in fact, searchable using most any word processor or web browser – saving researchers hours of looking up specific words or names.

An additional benefit is, once a primary source has been made public in some way –

1 http://www.domesdaybook.co.uk
2 http://www.nationalarchives.gov.uk/domesday/discover-domesday/

whether scanned or converted to a searchable format – it becomes a "safe" document. Fewer people will need to touch it, so it can be kept in proper storage and will not deteriorate. In addition, replicating a document electronically means that, even if the original is lost (say, to fire or theft), there will always be many other "copies" of that work. The original might be gone, but the content won't be.

In fact, technology can and does encourage electronic replication of works, in part because storing files electronically is more space- and cost-efficient than storing paper. Many organizations that house information, including historical documents, also wish to make it more readily available, and will work to replicate documents with the intention of sharing them (see, e.g., The Domesday Book, above).

In addition to making known documents more readily available, technology is increasing the number of primary sources available in the first place. A number of "crowd-sourcing" projects (some of which were created before that term came into common use) have increased the amount of material historians have to work with. For example, The Veterans' History Project, started by the U.S. Library of Congress in 2000, seeks letters, photographs and recorded interviews of veterans, to add to the collection.[3] Anyone (everyone) is invited to contribute, and in its first ten years, volunteers had donated over 65,000 *collections* of objects and recordings. These collections include objects that might have otherwise languished in attics, and recorded stories about war experiences that might never have been told.

Another way that technology is making historical research easier is the use of crowd-sourcing and other more intentional dedicated volunteer efforts to "translate" old hand-written materials into easy-to-read modern fonts. Software like ReCaptcha relies on the work of millions of people to transcribe old documents one letter at a time. Other more deliberate efforts have volunteers with practice in reading old texts transcribe one paragraph at a time. The Smithsonian Institution, for example, relies on approximately 6,000 volunteers to transcribe old hand-written letters into searchable text. So far, they have transcribed over 150,000 handwritten written pages from 1,000 projects.

Even more advanced technology can be used to authenticate historical documents, especially objects. Everything from an MRI scan to carbon-dating testing to chemical peels can help determine where an object is from, its contents, its age, whether it has been altered, and even whether it is a forgery. A very simple (and old) example is using a strong magnifying

[3] https://www.loc.gov/vets/about.html

glass to look for irregularities in a counterfeit dollar bill. In a more modern and complex example, museums can now use 3D technology to replicate broken pieces of statues and other artifacts. Preservationists can use the plastic 3D pieces to solve the "puzzle" of how the broken statue fits together without fear of harming the art. Once the piece's 3D components are put together properly, the actual historical piece can be rebuilt, following the 3D model's blueprint.

Technology is also helping historians share the results of the research as well. Most obviously, it is easier to type a paper using word processing software instead of a typewriter. But even document drafting is better than it was 20 years ago. Today, it is standard for writers to easily insert graphics into their works. New classes are being held to teach historians how to use graphing, statistical, and even map-creating software to aid to their research and publications. For example, the Institute of Historical Research of the University of London offers short graduate-level classes in Mapping and Geographical Information Systems, Designing Databases for Historical Research, and Data Preservation. Indeed, the IHR notes that sophisticated readers now *expect* that books and articles about history will include all of the above.

Limitations on Using Technology in Historical Research

Not all hardships faced by historians doing research have been eliminated through the use of technology, and indeed, technology can sometimes lead people to an erroneous conclusion without the intervention of a historian.

To start, the background knowledge of a historian is still necessary to interpret information from the past. Having access to data is not enough, without proper context. A historian must still put in the effort to learn the facts of his area of expertise, as well as trends and general observations of other historians. When the historian learns new facts, he must consider them in the context of what he already knows.

For example, imagine a novice researcher looking at arrest records in the American South between 1945 and 1965. Looking purely at data, the researcher would probably find that African-Americans were arrested at much higher rates than white Southerners. Without knowing the historical context, the uneducated researcher might conclude that African-Americans were more likely to commit crimes than whites. A historian familiar with that era and geographical location, however, would probably not reach the same conclusion, knowing

that pre-Civil Rights Act Southern states frequently arrested African-Americans for even minor infractions or for no crime at all, while not enforcing similar standards for whites, and failed to prosecute whites who committed crimes against African-Americans. In other words, the seasoned historian would know that the pure data does not tell the whole story. Instead, his background knowledge would allow him to conclude that the data is either misleading when taken alone or is support for the general proposition that African-Americans were not treated fairly under the justice system of the time.

Historians also still must rely on their training to spot information that is irrelevant, discredited, or outright fake. Judging relevance is a matter of experience. For example, a historian writing a biography of Emily Dickinson would understand that a letter written to the poet by a childhood friend would probably reveal something about her. But a letter written to Emily Dickinson announcing the opening of a new general store would not. Human are capable of reading the two letters and understanding the substantive difference. Technology might be capable of word-searching the two documents, but could not conclude that one was relevant to the poet's life, but the other was not.

Accurate translation also cannot be achieved by technology alone. In the context of modern translations from one to another, humans are still needed to identify and translate idioms and slang. When it comes to "translating" (ie, transcribing) old handwritten text, computers cannot take even the first step. Human help is necessary to – at a minimum – teach a computer what an old-fashioned "S" looks like. Beyond that, humans are still needed to modernize old languages, standardize old spellings, and convert old shorthand into longhand.

There are even ways that technology can hurt the study of history. One way is that modern communication allows for the faster spread of false information, as well as true information. Combine this with technology's ability to create believable and hard-to-spot forgeries, and researchers might be flooded with a lot of false information. This will not only interfere with their research and their ability to reach accurate conclusions, but their writings will have to explain why competing (and false) information is not reliable.

Related to this is the possibility that, in the age of "big data," historians may wrongly assume that lots of information about something reveals its importance. Imagine a novice historian 500 years from now, looking at events of the early 21st Century. Her sources are newspapers and an old social media (now long defunct) called "Twitter." Her research might

turn up 1,000 newspaper articles about some airplanes on September 11, 2001. And it might turn up 10,000 "tweets" about some failed music festival on April 28, 2017. Without context, this novice historian might conclude that the April 28, 2017 event was more important and more harmful than the September 11, 2001 event. To put it another way, "big data" is not the same as "all data" and it is certainly does not denote the importance of an event.

Conclusion

The last two decades have seen a significant increase in the ways technology furthers historical research. Sources are both easier to discover and easier to read. More material than ever before is available for general consumption, leading to both better historical research and less likelihood that historical documents will ever be completely lost or destroyed. Crowd-sourcing is both contributing to the amount of information available for research, and to some of the laborious steps needed to work with the information, such as translating it and transcribing it. The ability to search the text of a document decreases time spent on research and increases accuracy.

But the role of the historian is still essential in the study of history. Only an experienced and educated researcher can determine what information is both relevant and reliable. Only an experienced historian can take all of this newly-available information and assess it in its proper context. And only a historian can reach conclusions about the information before him. Although technology makes the job of actual research must easier than it was a mere 20 years ago, the intellectual rigor needed to study the information and present reasonable conclusions can be found only in human researchers, not computers.

Additional Resources

- American Historical Association, "Bag Data: An Opportunity for Historians?" available at https://www.historians.org/publications-and-directories/perspectives-on-history/march-2012/big-data-an-opportunity-for-historians
- Atlas Obscura, "How the Smithsonian Institution Is Crowdsourcing History" available at http://www.atlasobscura.com/articles/how-the-smithsonian-is-crowdsourcing-history
- Institute of Historical Research – various training classes, available at www.history.ac.uk/research-training/courses
- Times Higher Education, "History: the key to decoding big data" available at https://www.timeshighereducation.com/features/history-the-key-to-decoding-big-data/2016026.article