3 A method for navigating the infinite archive

William J. Turkel, Kevin Kee and Spencer Roberts

The transaction costs of traditional research methods

Although the conventions of historical research have evolved over the past century or so, until relatively recently most historians faced scarcity of information and limited access to sources. Most sources had to be consulted in person by visiting archives and libraries. Close reading, then as now, required a significant investment of human labour, as did the creation of indices, concordances, summaries, finding aids, and other scholarly apparatus that facilitated the discovery and use of new evidence for the past. Traditional scholarly conversations, at least those held in public, were conducted largely in journals and monographs, and unfolded over timelines typically measured in years.

As late as the 1980s and 1990s, two of us (Turkel and Kee) were taught the following method of research:

1. Do a systematic literature review.
2. Formulate question(s).
3. Do some combination of archival work, field work, or experimentation, depending on your discipline.¹
4. Master your subject.
5. Write up your results (for most historians, a well-defined cycle of conference papers and journal articles, leading eventually to a monograph).
6. Review the recent literature for any late-breaking news that needs to be taken into account.
7. Publish the final product.

When we learned this traditional research method it was already of questionable utility. We will argue that it is now more-or-less useless, made obsolete by the events of the last few decades. Sadly, however, it is still being taught in college and university departments that continue to act as if they are preparing students for academic life in the mid-twentieth century.

In the 1960s, Gordon E. Moore (later one of the founders of semiconductor giant Intel, but then director of research and development at Fairchild Semiconductor) observed that the number of electronic components on an integrated circuit was
doubling roughly every year and argued that ‘over the short term this rate can be expected to continue, if not to increase’. More than forty years later, ‘Moore’s Law’ continues to hold, having become a self-fulfilling prophecy for the industry. Computers and related electronic devices continue to become ever smaller, faster, more powerful and less expensive, digital storage becomes ever denser, and the number of devices connected to the internet continues to double at roughly the same pace. Digital sources, both those that were digitized from analogue originals and those ‘born digital’, also proliferate at exponential rates.

One gets a vague impression of this runaway growth by conducting life online and by participating in social media. Each day, the number of Google searches, Yahoo! searches, Bing searches, emails sent, text messages sent and YouTube videos watched, all number in the billions. We have a sense that the world online is big, probably in much the same way that a fish has the sense that the ocean is big: it seems unbounded in every direction. This local view, the view from a web browser, cannot really give us any idea of what is out there, however. In 2003, researchers at the University of California, Berkeley estimated that the amount of new information created the previous year was on the order of 5 exabytes ($10^{18}$ bytes). If one megabyte ($10^6$ bytes) is printed as text, it takes about as much physical space as a traditional book. Five exabytes-worth of traditional books would fill 37,000 libraries the size of the US Library of Congress. Seven years later, a similar study estimated that Americans consumed about 3.6 zettabytes ($10^{21}$ bytes) of information in 2008. Printed in traditional book form, this much information would blanket the United States and Alaska to a depth of about seven feet. A number that could be visualized in terms of library buildings at the turn of the millennium requires a continent to imagine less than a decade later. If you hated mathematics in school, this is what the word ‘exponential’ means. ‘When something grows exponentially, for a long time it may seem not to be changing at all’, Abelson, Ledeen and Lewis write in a study of the ‘digital explosion’. ‘If we don’t watch it steadily, it will seem as though something discontinuous and radical occurred while we weren’t looking.’

This glut has serious implications for the work of historians, as for every other kind of scholar. In a talk at Research Without Borders at Columbia University in 2011, Dan Cohen, Director of the Roy Rosenzweig Center for History and New Media at George Mason University, pointed out that a single historian might have been able to read and analyse the 40,000 memos issued at the White House during the Johnson administration, but that such a practice could never handle the 4 million email memos sent while Clinton was in office. Our digital libraries are growing as fast as our digital archives. Google Books, an effort to digitize every book published in modern history, has completed more than fifteen million volumes in the first decade of the project, and the speed at which they are working continues to increase. The company estimates that there are about 130 million books in the world; it is not unreasonable to speculate that they might finish digitizing them all within the lifetime of our students, if not ourselves. The traditional research method that we described above is clearly inadequate for dealing with an archive that is instantly accessible, machine-readable,
growing exponentially and constantly being reordered. Anyone who thinks that they are capable of doing a systematic literature review these days is mistaken, and anyone who thinks that they have mastered a subject is dangerously misguided. As one begins to learn about anything, the amount of new information on that subject will accumulate faster than it can be read or understood. The long-term projects that we have tended to favour as a discipline (especially the traditional monograph) are almost guaranteed to be way out-of-touch by the time that they reach print ... and still more so by the time that anyone has a chance to read or respond to them. What we need now are humility, curiosity and nimbleness; or to put it in the terms of Isaiah Berlin’s classic work, we all need to be (or become) foxes rather than hedgehogs.  

But how do we research effectively, efficiently, and as comprehensively as possible in the infinite archive? We have no wish to throw out the skills that lie at the core of our discipline: there is no substitute for close and critical reading, for careful citation, or for reasoned judgement. At the same time, there is no point in wasting human care or attention on tasks that can be done much faster and more thoroughly by machine. 

To date, humanists have shown little willingness to make use of new technologies. ‘The professoriate may be more liberal politically than the most latte-filled ZIP code in San Francisco’, Dan Cohen writes, 

but we are an extraordinarily conservative bunch when it comes to the progression and presentation of our own work. We have done far less than we should have by this point in imagining and enacting what academic work and communication might look like if it was digital first. 

Historical papers are mainly written with reference to previously published journal articles, books and chapters, most of which are either in print or else are digitized versions of the printed format. Most online journals imitate print versions without providing even rudimentary tools for social interaction. The situation is a lot like the televised radio shows that were produced at the birth of TV: new media, old mentality. The preferred format for publishing scholarly work, especially in history, continues to be printed journals or books. Though these media are perfectly acceptable, a reluctance to embrace alternatives such as web publishing will only hurt scholarship as years pass. 

The problem lies not only with the traditional establishments, but also with the scholars themselves. Basic word processors such as Microsoft Word are mainly used as digital typewriters, with few instructors or students making use of powerful cross-referencing abilities or even built-in reviewing tools. Researchers may use a computer to download digital versions of journals or books, or to search library catalogues, but often fall back on exploring the stacks themselves in order to find the marginalia that makes or breaks a project. Technology is often used to make the traditional methods a little bit easier without challenging standards or creating alternative procedures and tactics.
Organizations that serve scholarship have also found themselves on the trailing edge of technological trends. Only in the past couple of years have the Modern Language Association and the *Chicago Manual of Style* updated their standards to accommodate the use of websites, online publication, blogs and databases. A related problem is that traditional history has largely been based in text. While digital history makes use of text, too, there are any number of other media available, including hyperlinks, images, videos, audio recordings, visualizations and sonifications, equations, code, server logs, and still more exotic forms. In order to capitalize on these media, new and old, history needs to shift focus from text to more integrated means of conveying information. This is particularly important as society continues to move further from a text-based culture towards one that incorporates aural and visual displays. Humanists need to present their research and teaching in multiple channels simultaneously.

One real problem here is that lecturers and professors tend not to consider the learning and teaching of digital skills to be part of their mandate. The most foolish profess an asinine commitment to 'Luddism', but we have yet to meet anyone who knows what that word means and would also be willing to give up the technologies of reading, writing or publishing. Most also seem to be completely accepting of underwear, toothbrushes, electric lights, bicycles, buildings, microwave ovens and pharmaceuticals. Being a 'Luddite' in the academy turns out to be a weak and incoherent commitment. More charitably, many teachers subscribe to the myth of the 'digital native', assuming that their students are more technologically savvy than they are themselves. This, alas, turns out to be quite wide of the mark. Many of those teachers who do value digital skills assume that it should be someone else's responsibility to teach them, although whose responsibility is not at all clear. In Canada, at least, the vast majority of students majoring in history have never taken a computer science or programming class at the university level. If we do not incorporate these skills and techniques into our own classrooms, there is no reason to expect the current situation to change.

**A method for digital research**

There are any number of advanced computational techniques that can be used in humanities research, including text mining, machine learning, social network analysis and dynamic visualization. There are also sophisticated tools such as databases, web crawlers, search platforms and geographic information systems. Researchers who identify themselves as 'digital humanists' or 'digital historians', and many who do not, are already putting these tools and techniques to good use. We believe, however, that the work of all humanists can benefit from the adoption of a simpler, more fundamental digital workflow. The processes and tools that we will describe are not more complicated than those used for email or word processing, and the benefits obtain for any project, not only those that necessitate a high-tech approach. The seven fundamental principles of this method are:

1. Make everything digital.
2. Keep stuff in the cloud.
1. Make everything digital

If you have never really worked with digital sources or tools (besides word processing or exchanging email) there are a number of gentle introductions. We like Lisa Spiro’s *Getting Started in the Digital Humanities*, and the wiki of digital research tools which she edits. On the latter site you will find step-by-step instructions for many common tasks: ‘I want to ... analyze data, analyze texts, author an interactive work, blog, brainstorm/generate ideas ...’. The magic ingredient for any of these tasks is a digital source. Usually you can find digital sources easily; learning how to do advanced searching with the major search engines really pays off here. Of course, one always has to be on guard against the idea that if a source is not digital, it does not exist! When you cannot find a digital source, you should make one of your own. If you have to look at a traditional (analogue) source once, you should digitize it, so that your computer can help you to find it quickly if and when you need it again. Any article, photograph, map, letter or book can be digitized and made searchable through the use of Optical Character Recognition (OCR) software. Simply collecting digital sources and storing them on a hard drive is analogous to creating a library without a catalogue. If a source has been collected during research, it can and should be made accessible for all future purposes by taking a few extra minutes to run OCR on the text and add the metadata required to keep a digital database. The last thing you want to do is engage in long and fruitless searches for things that you know you read somewhere.

Software such as Adobe Acrobat Pro makes quick work of scanning, recognizing and adding metadata to a file that will allow your computer to retrieve the document when needed. Flatbed scanners are best used for books, larger media, and fragile objects such as photographs and slides. Printed documents can be easily scanned with an automatic document feeder that scans a stack of pages without user involvement. For short passages, handheld pen scanners are useful to trace over and capture the text, much the same way as a highlighting pen works. Photographs taken with a digital camera can also be recognized, and are particularly useful for capturing large or complex pieces. Improvements to camera phone technology make mobile devices a handy way to digitize a document and even the camera capabilities on the iPad are adequate for some documents, though you might have to play with lighting to ensure a clear picture. As a general rule, if a camera of any kind can take high-quality images of printed text, OCR software will be able to recognize it.

Adobe’s Portable Document Format (PDF) stores both an image of the page and, optionally, a layer of text that has been recognized with OCR. The former
is used by humans for reading, the latter by computers for searching. Additionally, Adobe Acrobat (or similar programs) will usually prompt the user to enter metadata about the document to facilitate searching and citation. Basic metadata such as author and title are crucial, but extended metadata help categorize and track documents. Some online content providers use OCR on their own documents to facilitate keyword searching, but strip out the text layer before providing it for download by the user. Acrobat Pro allows you to re-OCR the document, thus making it very much more useful.

Software such as DevonThink and Evernote allow you to create a library of digital sources that are organized by tags, categories or folders that are relevant for your own project. When searching for a term in DevonThink, for instance, a user can specify which folders to include in the search, expanding or limiting the scope of results. A user can also interact with digital sources through Acrobat or DevonThink in ways that are familiar from working with paper documents: adding comments and sticky notes, highlighting or underlining important passages, and the like.

2. Keep stuff in the cloud

One of the most important advantages of working solely with digital sources is that it is possible to duplicate your entire research project quickly and very inexpensively. Obviously you should do this for backup purposes: the time to create both local and offsite backups is before catastrophe strikes. If your research project is distributed across paper notebooks, physical books, photocopies, microfilm, fiche, 3x5 cards, and so on, you cannot really create a complete backup. More than one scholar has lost years of work and had to abandon a monograph or dissertation when his or her house or office caught fire, and practically everyone who has done a large project has worried about such things.

Digital backups have additional advantages beyond what librarians refer to as the LOCKSS principle (Lots of Copies Keep Stuff Safe). One is version control. Programs like Apple's Time Machine allow you not only to recover lost files, but to revert to earlier versions of a particular text. A less fancy mechanism, but one that is just as serviceable, is to duplicate a file each time you edit it and add the date to the filename in the YYYYMMDD format. Dates in this format will automatically sort in the correct order in both Windows File Explorer and Mac OS X Finder, making it easy to trace a history of revisions through filenames alone.

Digital files are also easily, inexpensively and privately stored on servers maintained by third parties and accessed via a web browser. Such ‘cloud’ storage gives you access to your research from any computer that has an internet connection. Some of the best-known and most useful examples are Google Docs, Dropbox, Evernote and JungleDisk. A service like Google Docs provides scaled-down office software in the browser (word processor, spreadsheet, etc.) while allowing documents to be shared with and modified by collaborators. Dropbox is integrated with your own system of files and folders, making sharing as easy as dragging and dropping. (We wrote this paper together by sharing Scrivener and
Word files in a Dropbox folder, for example.) If you've ever been frustrated trying to coordinate a co-authored article or grant application by emailing attachments to your colleagues, give one of these services a try. Other cloud-based programs, such as Evernote, are designed to provide one place to store your notes so that they can be accessed on every internet-enabled device that you own. JungleDisk provides automated, inexpensive online backup that can be scheduled to run unattended.

3. Manage citations in a database

The necessity for keeping good metadata is one of the first principles that novice scholars learn. The proper place for such information is in a database that is specifically designed for it. Citation management systems include freely available programs such as Zotero and Mendeley, both of which allow cloud-based storage and collaboration, and commercial alternatives such as Sente, Endnote and RefWorks. Each of these programs has advantages and disadvantages. Zotero, for instance, can scrape metadata from sites like Google Books, Google Scholar, Open WorldCat or Amazon.com, and will automatically store URLs and page images for sources found online so that you can access the version that you consulted even if the online version changes. Many of these tools can export citations into word processors using pre-formatted style guides, and can automatically create bibliographies from the selected sources.

4. The information comes to you

Think of a search as an activity that you do once, when you are looking for something in particular: a name, a date, a fact. When you find what you are looking for, you skim the results, make a note of the source, and, perhaps, bookmark the site if you think that it will be useful in the future. When you want to keep up with the news on a topic, or want a steady stream of results, it is usually better to pursue a strategy that Tara Calishain calls 'information trapping'. Major search engines like Google and Yahoo! provide mechanisms for creating RSS feeds from searches. When you subscribe to an RSS feed (for ‘RDF Site Summary’ or ‘Really Simple Syndication’), your computer is notified whenever something changes. By monitoring RSS feeds from searches using a program called a feed reader or feed aggregator, you get a steady stream of news about your topic, combined into a single report. You can use a desktop application like NetNewsWire on the Mac or an online service like Google Reader.

Additional techniques can make this strategy even more powerful. First of all, it is possible to use a service like Feed43 to create RSS feeds for any webpage. This has a bit more of a learning curve, but it allows you to monitor anything on the web. Second, Yahoo! Pipes provides a mechanism for combing RSS feeds with other kinds of computational processing. Again, there is a learning curve, but it is well worth it. Finally, other tools called crawlers, spiders or bots search the internet for specific content and download or index it. Crawlers can be coded
from scratch, but for those without programming skills, programs such as DevonAgent offer significant modularity and depth. The major benefit of these tools is that they can be set up and then left to run while you work on other tasks. 18

5. Attention is the scarcest resource

As the late Roy Rosenzweig argued, historians now face a culture of abundance. 19 The scarcest resource in your research process is always going to be your own time. Ideally, then, you only want to pay attention to those things which absolutely require it, and to hand everything else over to computer programs. Software tools help to focus attention through indexing, concordance, visualization, clustering and relating. DevonThink, for instance, can index documents in a collection to create a concordance. Terms can be sorted by weight or frequency, excluded, or used to compare documents. Proprietary algorithms analyze documents or passages to automatically cluster them by similarity, summarize them, or find other sources that are most closely related. Computationally sophisticated researchers can write their own programs for text mining or machine learning, but tools like DevonThink bring those abilities within reach of non-programmers.

Another way to highlight otherwise obscure connections is to use visualization tools such as Stéfan Sinclair’s Voyant or the IBM web service ManyEyes. Kee and Roberts ran a set of digital humanities syllabi through a visualization program to identify which authors were read most often and which university courses required students to read similar material. Such work can be done the hard way, of course, but digital tools expedite the process and allow the researcher to focus on questions and interpretation, rather than bean counting.

6. Work with others

Social media such as blogs and forums provide the glue that holds together large communities of people interested in any given topic, whether paid professionals or amateur enthusiasts. These communities provide a wealth of knowledge and expertise, opportunity to engage in public history, and a collective intelligence that can be leveraged to solve problems outside the scope of any single individual. As our colleague Shawn Graham discovered during his attempt to implement history learning through the commercial game Civilization IV, the most knowledgeable communities are often found in unlikely spaces. The CivFanatics Forum provided for Graham a wealth of knowledge about how gamers modify the parameters of the software to replicate historical events, but also demonstrated a depth of historical knowledge about the ancient world that surpassed the content of many university-level courses. 20 Alternate Reality Games (ARGs) demonstrate how communities working in unison can solve problems that otherwise defy solution. When Jane McGonigal helped to build an ARG for the release of Halo 2, she could not have envisioned that the
players would become so adept at solving the problems designed by the game team. Collective intelligence worked so well that ‘as the players invented smarter strategies and honed their coordination skills to meet these challenges, the designers were pushed to imagine future challenges even more difficult and confounding’. Though ARGs have not yet been used extensively for answering research questions, historians, such as our colleague Rob MacDougall, have begun to experiment with the capabilities of communities engaged in investigating historical events.

7. Share

Many of the communication barriers that prevented wide-scale collaboration have been eliminated in the digital age. When physical travel or paper publication costs limited opportunities for the public sharing of research results, they could only be presented intermittently. Content can now be shared much more regularly, and the costs of not doing so outweigh any benefit to sitting on results. Many scholars find that posting to research weblogs help them to clarify their own ideas, while putting them in touch with potential collaborators, informants, readers, would-be students, journalists and others. Communities that use sites like Twitter can instantly share job, grant and conference announcements, links to articles and news, information about databases or new repositories of sources, and the like. Research results published online in an open access venue reach the widest possible audience, and highlight the importance of post hoc peer review. For better or for worse, they also make scholarship much more interactive and dynamic, linking researchers to much wider communities of participation.

The method hits the road

Turkel and Kee were both on sabbatical in 2010–11, giving them time to experiment with processes of research and teaching. Looking for a faster way to research and write a monograph — the previous one took him seven years — Turkel found Steven Johnson’s and Chad Black’s articles on the Macintosh program DevonThink. He realized that by using all digital sources and combining a number of off-the-shelf programs (including DevonThink) he could work much more efficiently. As a result of using a collection of programs to find, harvest, cluster, excerpt and keep track of digital sources, Turkel’s second monograph took about ten months to research, write and submit to a publisher. When Kee found out about this, he decided to try teaching the method to graduate students (and Roberts was one of them). We briefly report on the results of that experiment here.

The method was introduced to the students in the context of an introductory course on the digital humanities. The course honoured convention, requiring that students think about digital humanities through definitions, theoretical problems and examples. It also required that students think with digital
humanities, exploring the ways in which computing could support their conventional research agendas (ranging, for example, from studies of fifteenth-century Norse manuscripts to eighteenth-century New England fashion), and transform their established research practices. For the students, digital humanities thus became the process of incorporating computing into their own humanistic practice.

Each week the students examined a research process, and then experimented with using a specific tool to make that process more efficient. To support their exploration they were given blog posts about digital methods, tutorials created by the makers of each tool, and in some cases supplemental tutorials written by Kee and Roberts. During the first week, they backed up their files, used Dropbox to access and exchange course-related material, and installed Zotero to manage their own citations (as well as access the references for the course). In addition, they accessed and contributed to course content via a Wiki (and began learning to use a simple mark-up language related to HTML).

Building on this foundation, they then turned their attention to social media. Each student created a blog, where he or she was required to comment on class readings, in-class discussion, Wiki discussion, the blog posts of classmates, and the blog and Twitter feeds followed by the student. Each also created his or her own Twitter feed, and set up an RSS aggregator such as Google Reader. During the third and fourth weeks, students were required to become conversant with HTML, and to do some very basic text analysis with Wordle and the Google Ngram Viewer. The fifth week provided a bit of a respite, during which the students used Adobe Acrobat Pro and learned about doing OCR (Optical Character Recognition). During the last half of the course, the focus shifted to research tools and the organization and analysis of research notes. The students worked with Google Advanced Search and DevonAgent, Evernote, DevonThink (or similar tool for the PC) and Scrivener. In the final week of their abbreviated introduction to this method they were encouraged to explore a software tool of their choice.

During the last two weeks of the course, the eleven students were asked to reflect on their use of the method in their blogs, and to comment specifically on the processes and tools that they found the most, and the least, helpful. According to many of the students, the use of these processes and tools was transformative. While a professor needs to approach students’ praise of an assignment with healthy scepticism, the level of engagement was obvious and genuine throughout the term. At the same time, to the considerable credit of the students, they were very frank with their criticism.

Some complaints were unavoidable. Windows users were understandably frustrated that tools such as DevonThink and Scrivener were only available on the Macintosh operating system. In some cases, students found themselves overwhelmed by a surfeit of new interfaces and possibilities. DevonThink, in particular, came under withering criticism for an inordinate number of options and uses. One student captured the feeling of several classmates when she recalled ‘looking at the screen and feeling lost’.
In some cases, students were overwhelmed by assignments that required them to rebuild their research practices at the same time that they were being challenged to rethink, in this and other courses, how they conceptualized the humanities. In the midst of this storm, some held fast to the familiar. ‘Once I have something which works why should I change it?’ one student asked. If the usefulness of a tool was not obvious or demonstrable, many understandably rejected it. The foray into the use of HTML, for example, elicited near-uniform dismay. This technology was viewed by many students to be the domain of computer programmers whose job it is to design and develop websites. ‘Don’t get me wrong’, one student remarked, ‘I would love to be technologically gifted, but I’m not’. Many students thought it would be best to focus on their own areas of competence or specialization, and to build collaborative relationships whenever they needed to create something outside their own abilities.

Not surprisingly, time-starved graduate students most appreciated processes and tools that enabled them to do that which they had been assigned (in their several courses) more efficiently. As one noted, ‘overall, I am in favour of practical tools which help me organize, sort, and speed up my research. ... I need organization, efficiency, and reliability.’ ‘With these tools I can get through the drudgy bits a lot quicker’, said another. Connecting with scholarly communities, doing research for specific projects, organizing sources, reading more efficiently, writing more effectively and managing citations were their highest priorities.

Social media tools brought them into contact with scholars with similar research interests. Twitter was especially helpful here. Many of the students took to Twitter grudgingly, viewing it as a space best left to celebrities and people who want to microblog what they ate for lunch. They quickly became converts after finding, to their considerable surprise, active researchers working in their own domains. ‘I began by just browsing twitter accounts that I could follow’, one medievalist in training wrote, ‘ByzantinePhil, Medievalists.net, Early Scottish History, etc. I then followed what they tweeted ... and boy, am I glad I took notice. Medievalists.net especially allowed me access to various articles on medieval topics. Want to learn about Icelandic horses or medicine during the Crusades? ... twitter [sic] has your fix. ... Twitter has offered me a great wealth of information.’ More often than not, students were happy to lurk; few commented on the relationships that they had formed via Twitter. In the majority of cases, the students were only a few weeks into their graduate careers (nine of the eleven were MA students), and were more comfortable playing the role of learner, rather than contributor, in these virtual research communities.

Once they became aware of new sources, students used Evernote to sort them. This cloud-based application makes it easier to manage incoming streams of sources because it can be accessed by laptop, smart phone or tablet ... any device with a web browser. As one noted, the ‘ability to just throw things into Evernote means I don’t forget/misplace/lose any valuable information I come across’. The task of determining which resources deserved attention was made easier with tools like Adobe Acrobat Pro and DevonThink. Several especially resourceful students also used the software to keep up with their assigned class
readings for every course. ‘Acrobat Pro helps with, and even improves, my read-

ings of assigned texts’, one student confessed, ‘which has been particularly useful
during those weekends when I have hundreds of pages to get through for
upcoming seminars and want to do a good job understanding them’. ‘Though
I hesitate to admit the following’, confessed another, ‘I can use the concordance
tool [of DevonThink] to see which words occur most frequently in a document,
filter them by word length, and very quickly get an idea of what the document
is about and what terms are significant to the author’.

Another key component of graduate studies in the humanities – the writing
of essays – was also reconsidered and, for many students, altered. The majority
adopted Scrivener, a new writing tool. For some, it was a friend: ‘it goes out of
its way to make my life easier’. ‘It allows me to build my paper as I always
should have’, remarked another, ‘in sections specifically related to topics and
parts of my argument and then look at the flow of my writing and seamlessly
rearrange the parts to fit together in a more cohesive fashion’. The only software
that rivalled Scrivener in terms of popularity was Zotero. By the end of the first
month, many students were using Zotero to handle their citations in all of their
courses. ‘As the semester is quickly winding down’, one student remarked, ‘the
bibliographies I’ve put together in Zotero have proved vital for me in terms of
my ability to complete all upcoming papers successfully and on time’. Returning
to the utility of social media for researchers, many students quickly came to
appreciate the benefits of Zotero Groups, which as one noted, ‘connects me to
bibliographies that I no longer have to spend hours in the library trying to find
(if indeed I find them all anyway). So here at the end of this course I can definitely
say that Zotero will continue to be a part of my life and work, both inside and
outside academia.’

Students had to use Zotero to access the sources for the course, and this
requirement may have made early converts. In most cases, however, the limits
of the academic calendar meant less time to experiment with method. As a
result, a fast learning curve was key for the adoption of particular tools. Evernote
required almost no preparation: within five minutes of installing the software,
students were using most of its affordances. The multiple options of other tools,
however, required a commitment that some were unwilling to make. At the same
time, several students acknowledged that their patience had been rewarded;
accidental discoveries (and last-moment conversions) were not uncommon.

According to their own testimony, a small percentage finished the course
relieved that the experiment was over. Others, however, found that their
workflows had been fundamentally altered. They were employing novel ways of
thinking and new kinds of reading, such as using software to ‘read’ papers that
had been assigned in their classes. One student moved beyond a piecemeal
adoption of certain tools to a complete commitment to the method, and especially
to DevonThink. ‘Quite simply’, he noted, ‘DevonThink has revolutionized the way
I research, organize, read, and study. I no longer print articles, but rather down-
load or scan them, run them through an OCR process, and add them to my DT
library. ... After only four months of use, my DT library has 400 unique items,
100,000 unique words, and over 2.6 million words total, all of which can be searched and related within groups of documents.' This tool enabled him to combine source materials in a way that had previously not been possible. 'I can merge multiple documents and perform a similar analysis, identifying the common terms among a selection of sources.' As a result, he was able to see connections that had before remained hidden, and ask questions that had not previously occurred to him. 'Without DevonThink, such work would be preventatively painstaking and tedious; with this tool, however, I find myself uncovering deeper connections and asking important questions because I can focus on interpretation and synthesis.'

Perhaps the most compelling reflection, however, came not in response to a specific piece of software. According to one student, 'there was something more important that I took away from the tools as a whole, not just with regards to the trajectory of an individual method'. Instead, she noted that

the fact that I was exposed to so many different types of methods ... and that some of them ... actually turned out to be useful without being scary is really important to me. Not only do I now have some new tools to use while I'm doing research ... I'm also more open-minded towards using them in the first place and really trying to engage with them, rather than brushing them off.

Helping these emerging humanists take a more experimental stance towards computing may be the most important outcome of this course.

In conclusion, we need to be clear. We do not think you should adopt the method that we presented in this chapter. By the time you read this, after all, many of our processes will have been refined and some of our tools superseded by better ones. Other people in our community will have adapted versions of the method that are more suitable for particular tasks, and a few generations of students will have helped us to figure out what is working and what is not. As with everything now, all is flux. What we do think is that any scholar or student can benefit from becoming more mindful about their method ... not as something that one acquires once and forgets about, but rather as something that one practises every day, making continuous small improvements over the course of a lifetime. Method is like historiography: when it stops changing, it is effectively dead. We need to keep changing, too.

Notes

1 Obviously history favours the former, although subfields like environmental history, the history of technology and public history often involve more hands-on work.
2 G. E. Moore, 'Cramming more components onto integrated circuits', Electronics, 38: 8 (19 April 1965), 114–17. This quote on p. 115.


13 See, for example, S. Vaidhyanathan, ‘Generational myth’, Chronicle of Higher Education (19 September 2008). Our own experience with teaching both undergraduate and graduate students is that they tend to be similarly unprepared for digital work from one year to the next.


16 Handwritten text continues to elude OCR. If you are not going to type a full transcription, add a handful of keywords to the metadata.


18 Many historians are (rightly) suspicious that automated searching often fails to turn up crucial evidence because present-day technology is insensitive to nuance, fails to understand various kinds of circumlocution, does not deal gracefully with ambiguity, synonymy, metaphorical usage, etc. All of these concerns are valid. The best way to address them is to become a more sophisticated user or creator of technology, rather
than abandoning it altogether. The time that you spend carefully reading one document is time that you do not spend scanning through others.


20 S. Graham, *Re-Playing History: The Year of the Four Emperors and Civilization IV*, Subject Centre for History, Classics, and Archaeology, The Higher Education Academy.


24 This was the so-called ‘super-secret monograph’. See W. J. Turkel, ‘On operating in stealth mode’ (21 February 2011), http://williamjturkel.net/2011/02/21/stealth-mode/ [accessed: 25 March 2012]. Note that not all of the sources used for this project were ‘born digital’. Many were digitized as part of the process of research and writing.


26 Turkel has had more luck introducing his graduate students to HTML, CSS and simple coding, although in a different context (interactive exhibit design). The utility of HTML is not nearly as evident in the research workflow that we describe here, although we would argue that some familiarity with markup languages is a crucial component of twenty-first-century literacy.

27 Turkel has had similar experiences assigning student blogging and tweeting as required coursework for public history graduate students: it takes them a while to get used to the idea that they are going to have to start ‘doing history’ in the public eye. They are usually much easier to convince than members of their ‘regular history’ graduate cohort, however.