

## Historians in Collaboration

# Writing The Historian's Macroscope in Public

Shawn Graham, Ian Milligan, and Scott Weingart, October 2014

We'd like a book about big data and history. Will you write it?" read the e-mail.

"Yes," we said, "but only if we can do it in public."

How do you write history when it is possible to download tens of thousands of historical documents in an afternoon? In our book, we explain how to do just that, by taking the reader through the process of building a digital macroscope to examine this "big data." It's not as difficult as one might think. But it's not easy, either. We set out to write this book in the summer of 2013 because we wanted to break down some of the barriers to entry for digital approaches to history. These barriers can seem insurmountably high, especially to students. Despite decades of humanities computing, the computer seems still no more than a glorified typewriter. In our own classes, we often meet with resistance to looking at historical materials with digital tools, or to writing history in forms that don't look like essays. "I didn't study history to have to deal with computers!" was an actual end-of-year comment from one of our students. We set out to rectify this, flipping the switch on our public draft of *The Historian's Macroscope: Big Digital History*, [themacroscope.org](http://themacroscope.org), in September 2013, in time for it to be used in class.

## What Does a Macroscope See?

We wanted to show how a digital history mind-set is allied with public, and collaborative, history. Whether the "digital" is digitizing resources, improving metadata, launching spiders to retrieve information, building online exhibits, or data-mining thousands of documents, digital history necessarily engages with others' codes and services. It changes what we can do, what we can see, and how we communicate. As an iterative endeavor, digital history is also often about failing in public, and learning from those failures. In this piece, we reflect on what it is like to write a book together, in full view of the public, and how new technology has made this process feasible.

Historians have long labored in conditions of cognitive scarcity: there are only so many documents one can read in a lifetime. We are now, however, in an era of larger and larger data sets, a situation that has been coupled with a corresponding computational ability to deal with them. We avoided a prescriptive definition of "big data," defining it instead as "simply more data that you could conceivably read yourself in a reasonable amount of time." Herein lies our idea of the "macroscope," then. It is an approach that reveals large-scale patterns that can be discovered when we collaborate with algorithms. Historians already do this (quietly) when they rely on digital finding aids and databases like Google Scholar; in *The Historian's Macroscope*, we make this reliance explicit.

A "macroscope" implies that research takes place in a workshop or laboratory, with an investigator who keeps a careful lab notebook. We believe that notebook should be

kept online, in public, so that other investigators can follow the same paths through the masses of available material, and potentially reach different conclusions. Digital tools do not require positivist assumptions that there is one and only one answer; rather, they require the awareness that algorithms and data sets provide their own perspectives, and our lab notebooks allow us to see those perspectives more clearly. Digital history utilizes the deformation of materials to change our perspective on history, rather than to justify a given story about the past. It's the difference between writing, "there appear to be patterns that suggest . . ." versus "statistics prove that . . ." For example, in one section of the book, we use a computational method known as topic modeling to make sense of the 8,000 biographies contained within the *Dictionary of Canadian Biography* ([www.biographi.ca/en/bio](http://www.biographi.ca/en/bio))—a large number by any count, compounded by the fact that some of them are nearly article length. To simplify, topic modeling identifies patterns of discourse. The topics found include both familiar patterns in Canadian history and also novel ones, such as new nodes of people, whose life stories emerge as key connective exemplars in Canadian historiography. The role of "liberalism" and its changing definitions in Canadian political life appears to be a major factor in Canadian historiography, when it is viewed through a microscope.

\*\*\* IMAGE REMOVED \*\*\*

*Robert Hooke's microscope from Scheme I of his 1665 Micrographia. On permanent display in The Evolution of the Microscope exhibit at the National Museum of Health and Medicine in Washington, DC.*

### Why Write It Online?

Given the speed with which the digital world changes, we shared our materials *before the book was finished* so that if we were wrong, we could improve. We also felt that given the openness of the field of digital history—from blogging professors, to shared data sets, to post-publication peer review—we would be well positioned to try something new, to try something *open*. Each of us is an active blogger: trying out ideas in public, making our "lab notebooks" open to inspection and criticism, and—most importantly—continually learning about new tools and methods from each other. It seemed like a no-brainer to continue in this spirit with a book project.

Yet, this was a frightening idea too. What if nobody commented—or worse, what if people hid behind anonymity to attack our project? What if someone stole our material

and scooped us? To say we had apprehensions about writing our book in public, with first drafts for all to see, would be an understatement. But when we began corresponding about cowriting this book, the idea of doing something different made sense. In this we were inspired by other online projects, especially Jack Dougherty and Kristen Nawrotzki's edited volume *Writing History in the Digital Age* (the finished book is available at [www.digitalculture.org/books/writing-history-in-the-digital-age/](http://www.digitalculture.org/books/writing-history-in-the-digital-age/), and the work-in-progress, complete with comments and edits, is at [writinghistory.trincoll.edu](http://writinghistory.trincoll.edu)).

### Head Decapitated

We then faced the problem of managing a three-way collaboration across North America. Despite using Google Docs successfully for the book proposal, we found ourselves drowning in the process of coordinating e-mail when we tried to scale up to a book. Track changes in Microsoft Word? A nightmare. We replaced Word with Scrivener, which focuses on the visual organization of one's thoughts. Versioning was still a problem. We turned to some online systems to help us manage our work, but soon discovered that Github did not play well with Scrivener: "Head Decapitated" came the cryptic error message. We ultimately settled on using Scrivener in a shared Dropbox folder as our writing environment. To keep from editing documents simultaneously we set up a private messaging service to alert each other whenever we were in a document. This back channel became an effective mode for continuous discussion. Once a section or group of ideas was fleshed out, we would post it publicly to the site. The site is Wordpress-powered with Commentpress installed, allowing readers to annotate or comment on individual paragraphs within the text.

### Rewarding Hiccups

Writing in public has turned out to be very rewarding, although there were hiccups. We released bits and pieces, here and there, on the website, in what seemed to be reasonably self-contained chunks. Predictably, readers found these chunks difficult to contextualize. Commenters, expecting a sequential narrative, would wonder why we didn't explain a term or provide background to some concept. We also weathered some pretty trenchant (but accurate) criticisms. It's one thing to receive these quietly in your office when a peer review arrives, but another to undergo the process in public. These comments, however, allowed us to significantly improve our content. They also became teachable moments in our classes, as our students could see us experiencing peer review and grappling with feedback. In student comments at the end of the fall 2013 term, one student wrote, "Seeing Dr. Graham present his draft work online, and reading what people wrote, made me feel better about my own writing and want[ing] to show it to others."

The rewards have far outweighed the difficulties. Imagine our delight at seeing our work included in graduate syllabuses, followed by a subsequent spike in traffic to the site, before the book has been formally published. At the University of Milwaukee, for instance, Professor Amanda Seligman had her class spend a week reading and leaving comments, allowing us to strengthen the final product for future classes. We continue to see the book draft assigned in graduate courses, and each visit helps us improve the content further. Now, only a year after beginning, we have incorporated the feedback and suggestions into the manuscript to be submitted to the press.

## Speed, Not Haste

It might seem like a speedy process, but it hasn't been hasty. The process of building *The Historian's Macroscope* has been a fun one, and we hope that our path can help other historians along their own routes to writing in public and experimenting with the big data that is history.

*Shawn Graham is an associate professor of digital humanities in the history department at Carleton University. Ian Milligan is an assistant professor of history at the University of Waterloo. Scott Weingart is a PhD student at Indiana University. The Historian's Macroscope should be published by Imperial College Press in the near future.*

### ***Getting Started with a Digital Toolkit***

Where can you start, right now, with digital tools to enhance your research? The Programming Historian ([programminghistorian.org](http://programminghistorian.org)), an ever growing collection of tutorials, is an excellent place to start. There is a tutorial on that website for working with MALLET, a tool for using topic models to understand thematic associations within your texts.

If you are working with students, Voyant ([voyant-tools.org](http://voyant-tools.org)) is a free, user-friendly online tool for exploring patterns in words or phrases in a text. It doesn't require any programming. A researcher can load texts into the system by pasting them in their entirety into the start box. Alternatively, one can paste in a series of URLs. Imagine having a digitized diary online—paste the URLs in chronological order into the start box, and soon you'll be able to see change over time!

And, of course, there's always the *Historian's Macroscope* draft—still developing—at [themacroscope.org](http://themacroscope.org). We're always happy to read comments and discuss ideas with readers.