Data, Humanities and the History of Medicine: New Pedagogical Approaches

The centrality of data and born-digital documents in contemporary medical care, public health and health policy means that the primary sources for future, and even present, medical historians will increasingly take on unprecedented digital forms. Historians of medicine – and indeed historians of virtually everything – will need to be trained in new digital tools and methods to work seamlessly between analogue and digital sources available to them.

Such new demands present an opportunity for innovation in history pedagogy as much as in research. A deliberate effort to integrate data analysis techniques in history courses not only broadens students' analytical capabilities, but reifies the values and skills central to humanities research, enabling students to apply robust critical analysis to their future work with data, irrespective of their careers. Medical history courses provide unique opportunities to illustrate the cultural, social and historical contexts required for understanding data, something that much of contemporary data analysis training unfortunately neglects.

Arguments for the need to integrate data analysis, if not quantitative methods, into humanities courses has risen steadily with the rise of the digital humanities over the last decade. While there are many good reasons to bring quantitative methods into history courses, several fundamental concerns remain: do historians really need to learn quantitative methods to teach history courses? Even for a historian with the ability and inclination, is there really adequate time to teach quantitative skills without being too superficial? Doesn't such effort detract from developing skills in qualitative analysis essential to humanities courses?

One approach that tries to skirt, but not solve, these particular issues is to displace the typical notion of data that resembles a spreadsheet featuring rows and columns of structured information. To engage with data of this sort entails other requisite skills, such as creating and structuring the data as well as ways of mathematically analysing and visualising it. There are entire courses devoted to these sorts of things, and they are not in history departments. But this is only one approach to teaching with data.

Alternatively, one can emphasise the qualitative character of data, and thus elide the mathematical connotations of data in favour of a more textual conceptualisation. What is historical data? Where do we find (or create) it? How do we engage with it as humanists in meaningful ways? There is no easier example of textual data than the metadata of a particular physical object. To take an obvious and familiar example, a book's metadata consists (partially) of its author, the publisher, date of publication, and so on.

In particular, the metadata associated with historical medical images provides a remarkable opportunity for students not only to learn about medical history via visual culture, but also to learn about data in generic terms. For instance, consider an image of a typical public health tuberculosis poster from the United States in the 1930s or 1940s, in the distinctive Works Progress Administration style. Of course, one can inquire about who created it, how it reflects the design sensibilities of its day and how it portrays conceptions of disease and approaches to public health.

But talking about only a single image has limited potential to increase data literacy; the power of data derives from sufficient quantity to generalise about a phenomenon. More interestingly, consider a web image search for TB posters. One instantly finds many dozens of posters from seemingly innumerable contexts spanning several decades. Again,

from a purely visual perspective, one can immediately compare different artistic styles and messages, and ask comparative questions about the set of images.

At the same time, an entirely different set of questions arises that pertains not to the images themselves, but to the reasons one sees these particular images in the first place. What websites are these images coming from? What kinds of archives or cultural heritage institutions collected these and made them available? Why do certain images appear so frequently? What might remain invisible? How representative are the search results compared to what one could find by physically digging around in archives? Because image metadata is often not embedded in digital objects themselves, an image's relevance to a image search is determined largely by the words on the webpage where the image appears. Therefore, these kinds of questions can be addressed, though often not fully answered, by examining such proximal data – the contexts in which the images appear online – and even more importantly, why that might be the case.

Comparative searches raise even more questions. For a particularly striking example, compare an image search of AIDS posters to an image search for Ebola posters. AIDS posters tend to be carefully designed with provocative imagery that ironically or jarringly interacts with the (usually sparse) text. Ebola posters tend to look like info-diagrams with virtually no compelling design component (many feature a cartoonish clip-art aesthetic), foregrounding sheer recitation of information. It is not the case that good graphic design has ceased to exist. To what extent is the discrepancy a function of the funders of the posters? Or the social construction of the diseases? Or the contexts (i.e. websites) of the images? Such questions, even when directed at web search results, bring historical methodology immediately into focus, provoking questions about archives, labelling, categorisation, search processes, and how all historical research, digital or not, is subject to the same kinds of constraints and biases.

While image search results are in part an artificial creation of a mysterious algorithm, they also are a culturally constructed representation of reality – both a physical reality (the images actually exist somewhere and have been digitised and cataloged in deliberate ways), and a virtual reality of how people find information online. Students are forced to confront explicitly the occult mechanisms that provide them with information, whether for historical research or contemporary medical advice. As with any good magic trick, much of the crucial effort is not in the execution but the framing. Deliberately keeping data at the forefront of the discussion is paramount, because explicitly relating the analytical investigation of metadata to data in general makes the skills more obviously exportable and transferrable.

The pedagogical approach described thus far, relying heavily on relatively recent digitisation efforts, encourages an engagement with digital archiving and publishing that has little precedent in traditional history courses. The same kind of opportunity now presents itself in academic publishing as well. Just as we (as teachers) can incorporate images into classrooms on larger scales, and thus raise important issues about digital literacy, we as scholars must do the same for our scholarship – and deliberately take up important challenges facing scholarly publishing. Perhaps it has been disappointing to read about images and their context without any illustrations. To be sure, it would have been far easier and more provocative to actually *show* the kinds of images I have described.

Most scholarly publishing venues have scarcely begun to take advantage of digital publishing to better support visual media, to say nothing of innovative digital scholarship. Yet I am not suggesting that we move scholarship online simply because a digital platform makes it easier to reproduce images easily available on the Web (which is not always true,

anyway). More importantly, scholars now have different kinds of images to show, often visualisations they must create to facilitate their research. Many digital projects that rely on digital methods like text mining, network analysis or mapping simply cannot be effectively discussed or evaluated without often copious illustrations. Can traditional publication forms and methods handle such imagery? How can scholars who require such illustrations situate their work in the scholarly ecosystem when forced to circumvent processes and venues (i.e. print) that provide necessary credentialing and legitimisation?

Returning to pedagogy and data for a moment, we can see that images serve well as objects of metadata inquiry because they are discrete and concrete objects. It is easy to imagine a poster in an archive or hallway; it is easy to talk about and visualise sets of images and how they differ from other kinds of images. But many of the same considerations about the nature of data (as discrete bits of information) and how to critically interrogate its provenance can also be explored with an emphasis on textual transmission and knowledge production. For this, there is perhaps no better tool than the nemesis of so many humanities professors: Wikipedia.

Why allot scarce class time to something as unreliable and sometimes egregiously wrong as a publicly and often anonymously edited encyclopedia? Why not use medical history scholarship itself? The reason is that published books and articles, especially academic ones, carry for most students an implicit aura of authenticity and accuracy. This seems to be true for academics, too, who still prefer print scholarship for similar reasons. Both a blessing and a curse, Wikipedia has a far greater transparency to it because contributors must explicitly document everything, knowing that no one will assume they have any expertise.

So where is the data, you may rightfully ask? As an encyclopedia, Wikipedia frowns upon complex subjective analysis; just the facts, please. So the facts – discrete statements about what happened or declarative interpretive statements – become 'givens': that is, data. Because online sources are easier to find and access than many scholarly ones (books especially), Wikipedia can and frequently does show flawed textual transmission at work in a strikingly effective way. Data gets propagated by well-meaning but often hasty and methodologically suspect research. Examining the digital traces of medical history articles provides a crucial complement to more traditional textual analysis as well as qualitative analysis. More importantly, whether using images or text, keeping data at the forefront of the discussion removes some of the foreignness, if not intimidation, of engaging with data.

Content aside, Wikipedia foregrounds important questions about the processes behind knowledge production in general. A volatile and often inexpert source such as Wikipedia almost makes an implicit argument for the stability and authority of print scholarship and the editorial processes that facilitate it. Yet again, parallel to how increasing our use of images in teaching may spill over into creating a more vibrant visual culture in scholarly publishing, engagement and experimentation with new editorial and publication workflows in the classroom can help us more effectively mobilise our own research.

Everyone knows that good scholarship does not necessarily *need* to be in print. But many scholars remain rightfully sceptical that processes for reviewing, vetting and revising scholarship work equally well for online venues as for more traditional print ones. And of course digital publications introduce new challenges in terms of ownership and sustainability. Clearly, future forms of scholarship present a largely unwanted burden to traditional editorial and publication practices. But these are reasons to improve and adapt existing editorial workflows, not to ignore new publication technologies and processes.

Scholarly publishing has already started down a path where it could end up fundamentally controlled by large for-profit publishing houses that care far more about profit margins than scholarly standards. The scholarly societies and journals that have been so instrumental in building and energising communities of practitioners, including *Medical History*, have an opportunity to strengthen those bonds by establishing themselves as leaders of scholarly communication, expanding both *what we can publish* and *who can access it* without sacrificing quality or rigour. This is not merely for the sake of data, images or new historical methods, but for the vitality and relevance of history itself.

Although I have focused on describing new ways of teaching about data (via images and text) for the benefit of students, I have deliberately juxtaposed pedagogy and publishing practices to highlight their interdependency and mutual influence. I hope it is clear that scholars themselves stand to benefit just as much. Whether in teaching or publishing, resistance to change stems not from mere habit or blind adherence to tradition, but from uncertainty and unfamiliarity with alternatives and their potential. Engaging with images and metadata in the classroom, as well as teaching about textual transmission and publishing models via a variety of digital publication platforms and editorial processes increases our own comfort with moving forward into uncharted publishing territory, in both form and process.

Frederick W. Gibbs
University of New Mexico, USA