

Cover Letter

Dear Professors Booth, Pasanek, and Committee,

I write to apply for the position of Assistant Professor of Digital Humanities, at the suggestion of Professor Pasanek. I am currently Assistant Professor of History at Northeastern University and core faculty in the digital humanities group, the NuLab for Texts, Maps, and Networks; I also hold an appointment as affiliate faculty in the Northeastern program in Information Design and Visualization. I earned my Ph.D. in History from Princeton University in November 2013.

My work uses new and creative methods to explore the tension between quantification and data on the one hand, and lived experience and humanistic practice on the other. I bring to this project a strong professional background in critical and historical methods coupled with expertise in machine learning, software architecture, and data visualization that allows me to interrogate massive datasets with the nuance and mindset of a historian.

I am currently engaged in two major projects that advance history and the digital humanities, one more scholarly and the other more infrastructural. The first, "Creating Data," investigates the origins of computational culture in the administrative workings and priorities of the late 19th century American state. This is a work of archival history that develops new forms of reading digital sources on three different 19th-century state institutions I have chosen for their historical importance and rich digital archives. The first is the emerging bureaucracies for collecting nautical weather data, first standardized by the Naval Observatory under Lt. Matthew Maury in the 1840s and 1850s, digitized onto punch cards as early as the 1920s, and currently maintained by the National Oceanic and Atmospheric Administration. The second is the US Census Bureau between 1870 and 1940, as it created new forms of information storage (including, most famously, the punch card itself) and of data visualization. The last is the Library of Congress, which became the country's most important taxonomizer of knowledge after moving out of the Capitol into the Jefferson building in 1897.

Each of the subjects for this project created large datasets which are still in use today; they are some of the oldest digital sources in existence. My work shows that in analyzing "big data," we need the same professional standards of rigor and competence historians demand when working on foreign language archives. By focusing on provenance, elision, and reanalysis on the digital archive, I bring alive the choices made by state administrators. I show the way that topics like the "frontier line" and the "center of population," were massaged, promoted, and distorted by the wishes and goals of researchers and bureaucracies. Methodologically, the project provides a badly-needed new framework for how humanists can use quantitative techniques to engage in humanistic data analysis. Portions of the initial blog posts sketching out the first stage of the project have been frequently assigned in graduate courses, and the visualizations have been circulated around the world and incorporated into museum exhibits. I plan scholarly publication of the full project that can keep these varied audiences.

"Bookworm," my second major project, is a software platform that I co-direct. Established digital humanities tools like Omeka allow museums to present their archives online using a set of customizable but manageable libraries. Bookworm does the same for raw texts, allowing a variety of quantitative analyses and data visualizations within the constraints of non-consumptive research. It scales to millions of documents of hundreds of thousands of words apiece. Bookworm grew out of my dissertation work and the Ngrams browser that my co-director, Erez Aiden, built with Google in 2010. It is now a rapidly growing platform in its own right, allowing a variety of types of analysis. It allows users to precisely define the particular library they wish to interrogate, while a plugin system makes it easy to integrate tools from machine learning such as named entity recognition and topic modeling. Installations have proven useful to scholars and have helped direct the interest of the general public to particular collections. Institutions including the Yale Libraries, the University of Chicago, and the US Department of State's Historian's office have built their own Bookworm installations to better understand, explore, and share their collections. I am currently part of major National Endowment for the Humanities implementation grant to deploy Book-

worm on the Hathi Trust, the world's largest non-profit library, where it will allow new forms of access and analysis to tens of millions of books.

My interest in digital methods grew out of issues of quantification that were central to my graduate training in American intellectual and cultural history. My dissertation, "Paying Attention: The psychological subject in advertising, education, and culture, 1890-1960," showed how techniques of measurement created by the "New Psychology" in the late 19th century laid the groundwork for the contemporary "attention crisis" as they were adopted and interpreted by practitioners in education and advertising through the early twentieth century. By tracing the origins of concepts like the "attention span" out of psychology departments into the educational institutions and radio research labs, I show how the resulting cultural understanding of attention kept a veneer of science while masking profoundly socially situated ends. The project brought me into the digital humanities as I downloaded hundreds of thousands of books from the Internet Archive to better trace the social and intellectual situation of new metaphors of mind like "focusing" and "concentrating" attention. I have continued some research in recently-opened archives relevant to this project with articles in mind.

My contributions to the digital humanities are interdisciplinary and extend well beyond this core set of projects. In particular, I have come to take the Internet seriously as a means of scholarly communication. My online work generally combines algorithmic sophistication with a wide audience. My work extending machine learning to humanistic textual analysis places me in a fruitful dialogue with scholars in English departments, in particular. For example, my work last month on tracing "plot arcs" in multidimensional spaces defined by topic models and segments of television shows was quickly adopted, applied, and reinterpreted by many of the leading practitioners of large-scale textual analysis. Posts on this blog are frequently featured in *Digital Humanities Now*, and have formed the kernel of two articles in the *Journal of Digital Humanities*.

I integrate public humanities into my research at every opportunity. My writings and data visualizations in the *Chronicle of Higher Education* on the changing place of the humanities in higher education helped shaped a national conversation about whether the humanities are in "crisis." My work on quantitative detection of anachronisms in popular television shows and movies has been featured in media from NPR to the *New Yorker*, and led to my employment as a historical consultant for historical dramas such as Showtime's *Masters of Sex*. Specific instances of the Bookworm browser I have built have been featured in numerous media; one served as a major part of *The Atlantic's* coverage of the 2015 State of the Union.

My undergraduate teaching areas are predominantly digital methods, the history of technology, and the United States since 1877. In teaching the US history survey from 1877 to 1945 last semester, I drew heavily on the history of capitalism and the history of American technology, either of which I would be pleased to offer as standalone courses. I am developing a history of computing course that would combine elements of history of science and critical code studies with a focus on the labor and cultural implications of computation in the United States and abroad. Perhaps my most rewarding teaching experience so far has been a first-year honors seminar on "The History of Big Data." In it, students learn about early modern information management practices (and conduct original archival research in the Boston area to find and report on a 19th century set of "data" collected by an individual, institution, or company); study the culture of management and control that birthed the computers of the server age; and then spend several weeks surveying the history of the Internet and personal computing since 1977, culminating in the corporate and government scandals of the last 5 years. As an instructor of graduate students, I can bring to UVA both an introductory digital history course and an advanced course in "Humanities Data Analysis" geared to serve the theoretical and technical needs of students in the English department as well as history. I can also teach graduate courses as needed in text analysis, humanities data visualization, and GIS; and in American intellectual and cultural history since 1877.

UVA would be the perfect home for me to continue this work. Bookworm would complement your existing expertise in bibliography, and as I transition from directly maintaining the code base myself to guiding developers, the experience of long-term project and grant management at IATH and the Scholar's

Lab will be invaluable. I have benefited from interdisciplinary connections at Harvard and Northeastern that I think UVA is well positioned to facilitate; and my combination of computational, large scale approaches with public scholarship should produce fruitful connections in the history department. Many thanks for your consideration.

Benjamin Schmidt