

## The Digital Humanities Initiative at Princeton University

David Mimno

I've been asked to describe the Princeton Digital Humanities Initiative (DHI), but I know that many of you are from other institutions, so I'm going to talk about it in the broader context of Digital Humanities. The Princeton DHI is a collaboration of many people from different parts of the university. As a postdoctoral fellow in the Computer Science department, I represent the more computational side. Others will have other perspectives, and that's one of the great things about this initiative.

If there is one message I'd like to convey in this talk, it is that the real goals of digital humanities are not that different from traditional humanities. The methods and tools that have become available recently are not part of standard practice in the humanities, so we need a label for them: thus, "digital" humanities. But as these new technologies become more familiar, they will fade into the background, leaving the real humanistic questions in the forefront. Take the example of communications technology. Most of us would describe my iPhone as a "cellphone", but increasingly people -- especially younger people -- think of it as simply a "phone". If we are successful, there will be no such thing as "digital humanities", just digital tools that are a natural and integral part of scholarship.

I'll start by describing my take on the digital humanities, then talk about what has happened so far at Princeton. Next I'll cover some of the models we've looked at, and end with the individual working groups and their goals.

### **Digital Humanities**

I've been involved in digital humanities for about ten years, not quite as long as the term itself has existed. I divide the work done into two phases. In phase I we make material accessible to computers by scanning pages, photographing objects, running optical character recognition, and other processes. In phase II we use algorithms and computational interfaces to learn about the material we digitized in phase I.

Most of the attention goes to phase I. There are skills and tools that people need to master. There are shelves of books that need to be processed. There are infinite details that must be considered, and getting those details wrong can mean the difference between useful and useless work. We could spend all our time and resources digitizing.

People think less about phase II. I would argue that if scholars start with the

goal of their analysis and fill in the technology needed to get there, keeping everything as simple as possible, we could make more progress.

Let's look at an example, the Perseus Digital Library. I bring it up both because I'm familiar with it, having led the development effort for the interface you see here, and also because it is often considered one of the most influential digital humanities resources.

The screenshot shows the Perseus Digital Library interface for the text "C. Julius Caesar, *Gallie War*". The page features a search bar at the top right with a search button and a search result for "Agamemnon". Below the search bar is a navigation menu with links for Home, Collections/Texts, Research, Grants, Open Source, About, and Help. A progress bar indicates the current position in the text. The main text area displays the beginning of Chapter 1, section 1, which describes the division of Gaul into three parts: the Belgae, the Aquitani, and the Celts. The text is annotated with hyperlinks for various geographical locations like Rhone, Garonne, Marne, and Seine. On the right side, there are several gray boxes providing additional information: "Notes" by J. B. Greenough, Benjamin L. D'Ooge, and M. Grant Daniell (1898); "Latin" text by T. Rice Holmes (1914); "Places (automatically extracted)" including Rhine, Garonne, Seine, Rhone, Marne, France, Aquitaine, and Spain; and "References" (47 total) listing commentary and cross-references. On the left side, there is a "Table of Contents" showing the structure of the book and chapters, and a "View text chunked by:" section with links for book, chapter, and section.

Here you see an English translation of Caesar's Gallic Wars. On the right are gray boxes representing related passages in other texts: the original Latin, a line-by-line commentary, and a collection of reference from other works to this passage. In the Latin text you can click on every word and get a morphological analysis and a link to a dictionary.

What is it about this interface that makes it useful? It's not the source material. Due to copyright law, every text that you see here is from before 1923, when computing machines were barely even a theoretical concept. What's different is the network of interrelations that we've represented through hyperlinks.

Digital humanities is a broad field. Perseus is just one example. You can find a list of projects at Princeton from the Princeton DHI site (<http://digitalhumanities.princeton.edu>). It goes from text-based projects like the Dante project to more traditional databases like the Index of Christian Art, to a project using image analysis and physical models of shattering to reconstruct Greek frescoes.

## **Digital Humanities at Princeton**

The Princeton DHI was founded about a year ago. There are more than a hundred people involved, but the core executive group includes people from English, OIT, the libraries, and computer science. It's important to point out that this was an organic collaboration, not a top-down initiative. All of us were working in this field without knowing each other and met almost by chance. That experience informs our goals: we want to build on existing strengths at Princeton by bringing people together.

We've been fortunate enough to receive substantial funding from the Council of the Humanities. But we've also received financial or in-kind support from more colleges, departments, and organizations than I have time to name. Our initial phase of building momentum has been successful. Now we have to decide what we think is the best way forward.

## **Models for DH Centers**

Princeton is relatively late to the digital humanities, but that gives us the advantage that we can evaluate existing models. We looked at a large number of similar programs at other schools. I'll show you a small selection, along with how they describe themselves.

Perhaps the best known is the Maryland Institute for Technology in the Humanities (MITH). They have a large physical space. Their home page emphasizes text and image analytics: tools and methodology. I was invited to visit last year to give one in a series of lunch talks about data mining.

Scholars' Lab at UVA also has space, but they emphasize more direct support of faculty and student research. Scholars' Lab has recently started a "Praxis" program that trains early-career graduate students in digital technologies.

One of the most interesting groups is at the University of Nebraska. They emphasize support for researchers, like Will Thomas who has been studying American railroads of the 19th century. They're growing fast by hiring scholars with expertise in computational tools.

We don't want to copy any one of these models exactly, but we have something to learn from all of them. Princeton has many existing areas of strength, both in humanities disciplines and in computing. We want to build the connections that will allow scholars to find the right collaborations. We expect that a lot of this work will happen through students, both graduate and undergraduate.

## **Working Groups**

I'll now discuss the organization of the Princeton DHI. Within the initiative, we have four working groups. Programming is in charge of planning events, like invited speakers, tutorials, and "hackathons". The curriculum committee is both collecting existing computationally oriented courses and identifying needs. Infrastructure and sustainability address two related goals: first, how to make the right connections between humanists, technologists, data, and hardware; and second, how to make the results of people's work last after the researchers themselves have moved on. Finally, we're looking for more funding.

## **Dangers and Opportunities**

I'd like to conclude by summarizing some of the lessons I've learned in digital humanities.

First, we need to remember the goal of digital humanities. We often think of the unit of work as the digital humanities project. These projects put a lot of work into some collection, and the result is often a pretty web page. The project becomes stale, and eventually dies. The corpus and the interface are important, but our goal should be the same as it always was: to learn about the past, to think about art or culture or history in a new way. If we lead with our questions and let the technology fill in the details, we will do better work.

Second, we should not be training "digital humanists" who sit between scholarship and technology. No one person will do everything. Collaboration between experts will become the model for the near future. We do not need interdisciplinarians, who are equally comfortable in multiple fields, but hireable in none. What we need is enough common language for people to work together, and enough social engineering to make connections happen.

Finally, we should not focus on technology for its own sake. The goal of digital humanities is to become invisible. With each year, the level of technological skill of students is rising. My children know all about smartphones, but will probably never see a card catalog. They will not understand how not to use technology. Training this generation to ask humanistic questions and to think critically about the past using digital tools may look much more like traditional scholarly education than we expect.