



Journal of Digital Humanities

VOL. 1 NO. 1 WINTER 2011

DANIEL J. COHEN, EDITOR

JOAN FRAGASZY TROYANO, EDITOR

SASHA HOFFMAN, ASSOCIATE EDITOR

JERI WIERINGA, ASSOCIATE EDITOR

DANIEL LUDINGTON, EDITORIAL ASST

NATALIA CECIRE, GUEST EDITOR

ISSN 2165-6673

CC-BY 3.0



ROY ROSENZWEIG CENTER FOR HISTORY AND NEW MEDIA

GEORGE MASON UNIVERSITY

A Community-Sourced Journal

We're pleased to present the inaugural issue of the *Journal of Digital Humanities*, which represents the best of the work that was posted online by the community of digital humanities scholars and practitioners in the final three months of 2011.

We wish to underline this notion of community. Indeed, this new journal is predicated on the idea that high-quality, peer-reviewed academic work can be sourced from, and vetted by, a mostly decentralized community of scholars rather than a centralized group of publishers. Nothing herein has been submitted to the *Journal of Digital Humanities*. Instead, as is now common in this emerging discipline, works were posted on the open web. They were then discovered and found worthy of merit by the community and by our team of editors.

The works in this issue were first highlighted on the [Digital Humanities Now](#) site and its related feeds. Besides taking the daily

pulse of the digital humanities community—important news and views that people are discussing—*Digital Humanities Now* serves, as newspapers do for history, as a rough draft of the *Journal of Digital Humanities*. Meritorious new works were linked to from *Digital Humanities Now*, thus receiving the attention and constructive criticism of the large and growing digital humanities audience—approaching a remarkable 4,000 subscribers as we write this. Through a variety of systems we continue to refine, we have been able to spot articles, blog posts, presentations, new sites and software, and other works that deserve a broader audience and commensurate credit.

Once highlighted as an “[Editors' Choice](#)” on *Digital Humanities Now*, works were eligible for inclusion in the *Journal of Digital Humanities*. By looking at a range of qualitative and quantitative measures of quality, from the kinds of responses a work engendered, to the breadth of the community who felt it was worth their time to examine a work, to close reading and analyses of merit by the editorial board and others, we were able to produce the final list of works. For the inaugural issue, more than 15,000 items published or shared by the digital humanities community last quarter were reviewed for *Digital Humanities Now*. Of these, 85 were selected as Editors' Choices, and from these 85 the ones that most influenced the community, as measured by interest, transmission, and response, have been selected for formal publication in the *Journal*. The digital humanities community participated further in the review process through [open peer review](#) of the pieces selected for the *Journal*. Authors selected for inclusion were given time to revise their work to answer criticisms and suggestions from the community and editors, prior to a round of careful editing to avoid typographical errors and other minor mistakes.

To be sure, much worthy content had to be left out. But unlike a closed-review journal it is easy to see what we had to choose from, since the trail of Editors' Choices remains on *Digital Humanities Now*.

Inclusion in this issue is in many respects harder and rarer than inclusion in a print or print-like journal, since it represents a tiny minority (less than one percent) of the work that digital humanities scholars made public in this period. We hope and expect that this selectivity will reinforce the value of the work included.

Even with these several layers of winnowing, the result is a sizable and wide-ranging first issue, roughly 150 pages and four hours of multimedia. The most-engaged article of the quarter was by Natalia Cecire, whose post on theory in digital humanities sparked an energetic debate and many additional posts by those who agreed or disagreed. In response, we asked Natalia to be a guest editor of a special section in this issue on the topic of her piece, which she has introduced and knitted together with responses addressing digital humanities' awkward relationship to theory (or the lack thereof).

Beyond this special section, we have a slate of individual articles, including lengthy treatments of text mining and visualization, critical discourse and academic writing, the use and analysis of visual evidence, and a series of podcasts on humanities in a digital age. To start the issue, we have included a piece by Lisa Spiro on how to get started in digital humanities, and in what we believe is a first for the field, we end the issue with an entire section devoted to a critical engagement with tools and software.

We believe the variety of content in the *Journal of Digital Humanities* truly parallels the scope of work being done in the community. Because this journal is digital-first, we are able to take into account the full array of works produced in the discipline. Unlike other publications, we can, for instance, point to and review software, and we can include audio and video. We can also accept works of any length. We plan to maintain this emphasis, that there is no real or implied

pressure to submit a standard essay of 5,000-10,000 words or to flatten nonlinear digital works into a print-oriented linear narrative.

Our community- and web-sourced method has several other advantages over the traditional journal model. First, as we have already noted, many more eyes have looked at the content within this volume, ranging from perhaps superficial readers—hundreds who saw and read it in their RSS readers or via social networks—to more in-depth engagements, such as those who responded in comments on the site of the original work, wrote a response on their own site, or who participated in our open review of the selected works on the *Digital Humanities Now* website.

Moreover, we believe this model has helpfully led to the inclusion of contributors from a wide range of stations compared to a traditional academic journal. Represented in this volume are up-and-coming graduate students already doing innovative and important work, non-academics and technologists who focus on thorny and often intellectual questions of implementation and use, and those in fields that border on or intersect with digital humanities, such as librarians, archivists, and museum professionals. We believe this is healthy for the ideas and practice of the digital humanities community, moving it beyond an insular community of mostly tenure-track academic scholars.

In that spirit of inclusion, we hope that you'll join us in contributing to the *Journal of Digital Humanities*, as someone who finds and validates new work—as a daily editor on *Digital Humanities Now* or as a quarterly editor on the journal—or, like those whose work appears in this first issue, as someone who contributes greatly to the field by openly posting their work online.

Daniel J. Cohen and Joan Fragaszy Troyano, Editors

Getting Started in Digital Humanities

When I presented at the Great Lakes College Association's [New Directions](#) workshop on digital humanities (DH) in October, I tried to answer the question "Why digital humanities?" But I discovered that an equally important question is "How do you do digital humanities?" Although participants seemed to be excited about the potential of digital humanities, some weren't sure how to get started and where to go for support and training. Building on the [slides](#) I presented at the workshop, I'd like to offer some ideas for how a newcomer might get acquainted with the community and dive into digital humanities work. I should emphasize that many in the digital humanities community are to some extent self-taught and/or gained their knowledge through work on projects rather than through formal training. In my view, what's most important is being open-minded, experimental, and playful, as well as grounding your learning in a specific project and finding insightful people with whom you can discuss your work.

Determine what goals or questions motivate you

As with any project, a research question, intellectual passion, or pedagogical goal should drive your work. Digital humanities is not technology for the sake of technology. It can encompass a wide range of work, such as building digital collections, constructing geo-temporal visualizations, analyzing large collections of data, creating 3D models, re-imagining scholarly communication, facilitating participatory scholarship, developing theoretical approaches to the artifacts of digital culture, practicing innovative digital pedagogy, and more.

Get acquainted with digital humanities

- The [CUNY Digital Humanities Resource Guide](#), which was produced collaboratively, offers an excellent introduction to digital humanities, covering sample projects, syllabi, "hot topics," journals, and more.
- Nebraska's [Digital History Project](#) offers "a number of essays, interviews, and lectures by practicing digital historians along with a Directory of Digital Historians, project reviews, tool reviews, and digital history course syllabi, undergraduate class projects and graduate student digital research projects." (Hat tip Doug Seefeldt.)
- Ask or answer a question on [DH Questions & Answers](#), "a community-based Q&A board" where people weigh on everything from designing a digital history curriculum to computational analysis of perspective in art.
- Look through Blackwell's [Companion to the Digital Humanities](#), which collects essays from some leading thinkers in the digital humanities (and based on my preliminary research seems to be the most commonly assigned text in digital humanities courses). Another frequently assigned text is Dan Cohen and Roy

Rosenzweig's [Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web](#).

- Look through journals in the digital humanities, including [Digital Humanities Quarterly](#), [LLC](#), and [Digital Studies / Le champ numérique](#).

Participate in the digital humanities community

Frankly, I think that the energy, creativity, and collegiality of the digital humanities community offer powerful reasons to become a digital humanist.

- Attend a [THATCamp](#). At a THATCamp, participants spend the first session setting up the agenda, drawing from blog posts they contributed prior to the event. They devote the rest of the time to hands-on workshops and discussions about topics such as pedagogy, text visualization, and collaboration. These inexpensive, interactive, non-hierarchical unconferences typically are organized regionally (Toronto, New England, Bay Area, New Mexico) or by theme (pedagogy, publishing, liberal arts colleges, games, museums). You can learn a lot just by browsing the session proposals and summaries from past [THATCamps](#), so I'm excited that [PressForward](#), the innovative publishing venture from the Roy Rosenzweig Center for History and New Media, will soon publish [Proceedings of THATCamp](#). (I enjoyed my THATCamp experience so much that I worked with Andrew Torget and Anita Riley Dryden to organize [THATCamp Texas](#).)
- Go to a [digital humanities conference](#). The annual [Digital Humanities](#) conference sponsored by the [Alliance of Digital Humanities Organizations](#) (ADHO), features the latest research and a lively community. Disciplinary conferences such as the [Modern](#)

[Language Association](#) and [American Historical Association](#) include a number of digital humanities-related sessions.

- Participate in—or start up—a regional group, such as [Decoding Digital Humanities](#) (with chapters in London, Melbourne, Bloomington, and Lisbon), [Toronto Digital Scholarship \(DISC\)](#), [Digital Humanities in Boston & Beyond](#), or [Digital Humanities Southern California](#).
- Support a professional organization in digital humanities, such as the [Association for Computers and the Humanities](#) (I'm privileged to serve on the Executive Council) or the [Association for Literary and Linguistic Computing](#).
- Participate in an online community, such as the [Digital Americanists](#) (I'm one!), [Digital Classicists](#), [Digital Medievalists](#), [HASTAC](#), [EighteenthCentury.org](#), and [Romantic Circles](#).
- Take part in crowdsourcing projects that engage the public in contributing to scholarly work, such as [Transcribe Bentham](#).
- Review work in the digital humanities, such as by participating in an open peer review process (e.g. for the [Journal of Digital Humanities](#)) or by signing on to be a reviewer for a grant program, journal, or conference.
- Follow and interact with digital humanities folks on [Twitter](#). Not only is Twitter a great way to keep tabs on what's going on in the community, but it can help connect you with people, so when you meet them for the first time at a conference you already feel that you sort of know them.
- Read and respond to [blogs](#).

Stay informed

- I always learn something from [ProfHacker](#), a fantastic group blog focused on teaching, tools, and productivity. (By the way, ProfHacker was hatched at a THATCamp.)
- [GradHacker](#) covers software reviews, discussions of professional issues, and more, from a grad student perspective. (Hat tip Ethan Watrall.)
- Subscribe to the [Humanist Discussion Group](#), which is expertly facilitated by Willard McCarty and has supported conversation and information sharing since 1987.
- Check out [Digital Humanities Now](#), which brings together current discussions and news in the digital humanities community "through a process of aggregation, discovery, curation, and review."
- Follow what people are bookmarking on [Diigo](#) or [Delicious](#). (I'm a compulsive [bookmarker](#), but not so good about annotating what I come across.)
- Join the [Digital Humanities Zotero group](#), which collects resources on the digital humanities. (Hat tip Mark Sample.)
- Explore what's going on at digital humanities centers. Check out [CenterNet](#), an "international network of digital humanities centers."
- Connect with local digital humanities centers. For example, in the Great Lakes region, Michigan State University's [MATRIX](#) builds digital collections, hosts H-Net, offers training, and more. (Hat tip Ethan Watrall).

Explore examples for inspiration and models

To find projects, see, for example,

- NEH Office of Digital Humanities' [library of funded projects](#), browsable by type of project (e.g. toolbuilding, public projects, education).
- [NINES](#) lists peer reviewed projects in 19th century studies; [18thConnect](#) is doing the same for 18th century studies.
- [Arts & Humanities.net](#) allows you to browse projects by discipline, funding body, and content type.
- [The Institute for Enabling Geospatial Scholarship's](#) catalog of [Spatial Humanities Projects](#).
- Patrick Sahle's Catalog of [Scholarly Digital Editions](#).

Pursue training

Workshops and Institutes

- The [Digital Humanities Summer Institute](#), hosted at the University of Victoria, has an [excellent reputation](#) and offers week-long workshops on topics such as text encoding, multimedia, Geographical Information Systems, project management, and digital pedagogy, taught by leaders in the field. [Scholarships](#) are available, and the ACH provides travel bursaries for graduate students.
- The [Nebraska Digital Workshop](#), sponsored by the Center for Digital Research in the Humanities (CDRH) at the University of Nebraska–Lincoln (UNL), enables a select group of early career scholars to present their work to and get feedback from senior scholars.

- [NEH Institutes](#) explore key topics in the digital humanities and often cover travel costs. Upcoming institutes focus on TEI, spatial humanities, linked ancient world data, working with digital text, digital cultural mapping, and high performance computing.
- [The NINES Summer Workshop](#) offers in-depth training to scholars in 19th C British and American literature.
- Oxford University offers a week-long [summer workshop](#) in digital humanities.
- The University of Virginia's [Rare Book School](#) often includes sessions of interest to digital humanists, such as "[Born Digital Materials: Theory & Practice](#)," "[XML in Action: Creating Text Encoding Initiative \(TEI\) Texts](#)" and "[Digitizing the Historical Record](#)."
- The Women Writers Project (WWP), which has significant expertise in the Text Encoding Initiative (TEI), provides [seminars and workshops](#).
- The [Michigan State University Cultural Heritage Informatics \(CHI\) Fieldschool](#) brings together students to collaborate on cultural heritage projects, developing skills in project management, design, programming, etc. (Hat tip Ethan Watrall.)

Online tutorials

- Overviews: Stanford's [Tooling Up for Digital Humanities](#) provides helpful introductions to digitization, pedagogy, data visualization, and more.
- Doing online research: William J. Turkel's [Going Digital](#).
- Geographical Information Systems: the [Institute for Enabling Geospatial Scholarship's](#) tutorials on [Spatial Humanities Step by Step](#), UCLA's [GIS](#) tutorials.

- Text Analysis: [TAPoR Portal Recipes](#) (I'm a fan of [TAPoR](#), which provides tools for analyzing and visualizing texts, and a member of the TAPoR advisory group.)
- Text Encoding Initiative/ XML: [TEI by Example](#); the WWP's [Resources for Teaching and Learning about Text Encoding](#); Laura Mandell, Brian Pytlik-Zillig, Syd Bauman, et al, [XSLT-for-Humanists](#); John Bradley, Elena Pierazzo and Paul Spence, [An XSLT Tutorial](#).
- Programming: William J. Turkel and Alan MacEachern, [The Programming Historian](#); Jason Heppler, [The Rubyist Historian](#).

Learn standards and best practices

If you want your project to have credibility and to endure, it's best to adhere to standards and best practices. By talking to experts, you can develop a quick sense of the standards relevant to your project. You may also wish to consult:

- Arts & Humanities Data Service's [Guides to Good Practice](#)
- The [NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials](#)
- Peer review standards for [NINES](#) and [18thConnect](#)
- Grant guidelines (e.g. from the [NEH](#)) for technical standards, particularly regarding data management and sustainability

Find collaborators

Most digital humanities projects depend—and thrive—on collaboration, since they typically require a diversity of skills, benefit from a variety of perspectives, and involve a lot of work.

- [Digital Humanities Commons](#) serves as an online hub (or matchmaking service) where people can identify projects to collaborate with and projects can discover collaborators. (I'm a member of the advisory board.)
- Talk with library and IT staff at your own institution. Although many library and IT professionals are necessarily focused on the day-to-day, there is also an increasing recognition that what will distinguish libraries and IT groups is their ability to collaborate with scholars and teachers in support of the academic mission. Be a true collaborator—don't just expect technical (or content) experts to do your bidding, but engage in conversation, shape a common vision, and learn from each other. (Steve Ramsay offers great advice to collaborators in "[Care of the Soul](#)," and the Off the Tracks Workshop devised a useful "[Collaborators' Bill of Rights](#).") If you can bring seed funding or administrative backing to a project, that might make it easier to attract collaborators or garner technical support.
- Reach out to others in your community. By attending a THATCamp or corresponding with someone who shares your interests, you may discover people who can contribute to your project or help shape a common vision. You could also find a colleague in computer science, statistics or another field who has common research interests and would be eager to collaborate. You might be able to hire (or barter with) consultants to help out with technical tasks or provide project advice; I understand that Texas A&M's Initiative for Digital Humanities, Media, and Culture is exploring offering consulting services in the future to help advance the digital humanities community.
- Engage students. While there can be risks (after all, students graduate), students can bring energy and skills to your project.

Moreover, working on digital humanities projects can give them vital technical, project management, and collaborative skills.

- Consider a DIY approach. As Mark Tebeau of [Cleveland Historical](#) wisely observed at the New Directions workshop, if your institution doesn't provide the support you need for your DH project, why not strike out on your own? As Trevor Owens suggests in "[The digital humanities as the DIY humanities](#)," it takes a certain scrappiness to get things done in digital humanities, whether that's learning how to code or figuring out how to set up a server. If you don't think you have the time or skills to, say, run your own web server, consider a hosted solution such as [Omeka](#). In the long term, it's a good idea to affiliate with an institution that can help to develop and sustain your project, but you may be able to get moving more quickly and demonstrate the value of your idea by starting out on your own.

Plan a pilot project

Rather than getting overwhelmed by trying to do everything at once, take a modular approach. At the New Directions workshop Katie Holt explained how she is building her [Bahian History Project](#) in parts, beginning with a database of the 1835 census for Santiago do Iguape parish in Brazil and moving into visualizations, maps, and more. This approach is consistent with the "permanent beta" status of many Internet projects. Showing how a project moves from research question to landscape review to prototype to integration into pedagogy, Janet Simons and Angel Nieves of Hamilton's [Digital Humanities Initiative](#) demonstrated a handy [workflow and support model](#) for digital projects at the workshop.

Where possible, adopt/adapt existing tools

Explore open source software. Too often projects re-invent the wheel rather than adopting or adapting existing tools.

- Find tools via Digital Research Tools ([DiRT](#)) wiki (which I founded. Bamboo DiRT now has a new home and provides better browsing and sharing capabilities, thanks to the hard work of the fabulous Quinn Dombrowski and Bamboo).
- [SHANTI's UVa Knowledge Base](#) offers useful information about technologies, teaching, and research approaches. (Aimed at the University of Virginia, but more widely applicable.)
- You can also poke around GitHub, which hosts code, to identify tools under development by members of the digital humanities community such as [CHNM](#) and [MITH](#).

NITLE Can Help

Let me end with a plug for [NITLE \(the National Institute for Technology in Liberal Education\)](#), my (relatively) new employer. One of the reasons I wanted to join the organization as the director of NITLE Labs is because I was impressed by its [digital humanities](#) initiative, which my colleague Rebecca Frost Davis leads. Among NITLE's activities in the digital humanities:

- Hosting a (free) online [Digital Scholarship Seminar Series](#). Archived sessions include [Joining the National Digital Humanities Conversation: Communities, Conferences, Centers](#) and [Digital Scholarship in the Online Archive](#).
- Producing white papers, book chapters, and blog posts focused on digital humanities at liberal arts colleges, including

- ▶ Rebecca Frost Davis, [How to engage in digital humanities at small liberal arts colleges?](#)
- ▶ Rebecca Frost Davis and Quinn Dombrowski, "[Divided and Conquered: How Multivarious Isolation is Suppressing Digital Humanities Research](#)" (PDF)
- Offering (through collaboration between Rebecca and [Kathryn Tomasek](#) of Wheaton College) a workshop on "[Integrating Digital Humanities Projects into the Undergraduate Curriculum](#)."

If you're a veteran digital humanist, how did *you* get started, and what do you wish you knew from the beginning? If you're a newcomer, what do you want to know? What worries you, and what excites you? What did I leave out of this overview? I welcome comments [on my blog](#).

Originally published by Lisa Spiro on [October 14, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Articles

Academic History Writing and its Disconnects

Tim Hitchcock 2

Defining Data for Humanists: Text, Artifact, Information or Evidence?

Trevor Owens 6

Demystifying Networks, Parts I & II

Scott Weingart 9

Clustering with Compression for the Historian

Chad Black 22

Spatializing Photographic Archives

Marc Downie and Paul Kaiser 30

“Humanities in a Digital Age” Symposium Podcasts

Jeremy Boggs, Alison Booth, Daniel J. Cohen, Mitchell S. Green, Ann Houston, and Stephen Ramsay 32

Philosophical Leadership Needed for the Future: Digital Humanities Scholars in Museums

Nik Honeysett, and Michael Edson 33

Critical Discourse in Digital Humanities

Fred Gibbs 34

TIM HITCHCOCK

Academic History Writing and its Disconnects

The last ten years have seen the development of what looks like a coherent format for the publication of inherited texts online – in particular, ‘books’. The project of putting billions of words of keyword searchable text is now nearing completion (at least in a Western context); and the hard intellectual work that went into this project is now done. We are within sight of that moment when all printed text produced between 1455 and 1923 (when US copyright provisions mean that the needs of modern corporations and IP owners outweigh those of simple scholarship), will be available online to search and to read. The vast majority of this digital text is currently configured to pretend to be made up of ‘books’ and other print artifacts. But, of course, it is not books. At some level it is just text – the difference between one book and the next is a single line of metadata. The hard leather covers that traditionally divided one group of words from another are gone; and while scholars continue to pretend to be reading books, even when seated comfortably in front of their office computer, this is a charade. Modern humanities scholarship is a direct engagement with a deracinated, Google-ised, Wikipedia-ised, electronic text.

For the historian, this development has two significant repercussions. First, the evolution of new forms of delivery and analysis of inherited text problematizes and historicizes the notion of the book as an object, and as a technology. And second, in the process problematizing the ‘book’, it also impacts the discipline of history as it is practiced in the digital present. Because history has been organised to be written from ‘books’, found in hard copy libraries, the transformation of books to texts forces us to question the methodologies of modern history writing.

In other words, the book as a technology for packaging and delivery, storing, and finding text is now redundant. The underpinning mechanics that determined its shape and form are as antiquated as moveable type. And in the process of moving beyond the book, we have also abandoned the whole post-enlightenment infrastructure of libraries and card catalogues (or even OPACS), of concordances, and indexes, and tables of contents. They are all built around the book, and the book is dead.

To many this will appear mere overstatement; just another apocalyptic pronouncement of radical change of the sort digital humanists specialize in. And there is no question but that ‘books’ will continue to be published for the foreseeable future. Just as manuscripts continued to be written through all the centuries of the book, so the hard copy volume will survive the development of the online and the digital. But, the transition is nevertheless important and transformational; and for a start allows us to interrogate the ‘history of the book’ in new ways.

First, it allows us to begin to escape the intellectual shackles that the book as a form of delivery imposed upon us. That chapters still tend to come out at a length just suited to a quire of paper, is a commonplace instance of a wider phenomenon. If we can escape the self-delusion that we are reading ‘books’, the development of the infinite archive,

and the creation of a new technology of distribution allows us to move beyond the linear and episodic structures the book demands, to something different and more complex. It also allows us to more effectively view the book as an historical artifact and now redundant form of controlling technology. The 'book' is newly available for analysis.

The absence of books makes their study more important, more innovative, and more interesting. It also makes their study much more relevant to the present – a present in which we are confronted by a new, but equally controlling and limiting technology for transmitting ideas. By mentally escaping the 'book' as a normal form and format, scholars can see it more clearly for what it was. To this extent, the death of the book is a liberating thing – the fascist authority of the format is beaten.

At the same time we are confronted by a profound intellectual challenge that addresses the very nature of the historical discipline. This transition from the 'book' to something new fundamentally undercuts what historians do more generally. When one starts to unpick the nature of the historical discipline it is tied up with the technologies of the printed page and the book in ways that are powerful and determining. Footnotes, post-Rankean cross referencing, and the practises of textual analysis are embedded within the technology of the book, and its library.

Equally, the technology of authority – all the visual and textual clues that separate a Cambridge University Press monograph from the irresponsible musings of a know-nothing prose merchant – are slipping away. At the same time, the currency of professional identity – the titles, positions, and honorifics – built again on the supposedly secure foundations of book publishing – seems ever more debased. The question becomes: is history, like the book – particularly in its post-

Rankean, professional, and academic form – dead? Are we losing the distinctive disciplinary character that allows us to think beyond the surface and makes possible complex analyses that transcend mere cleverness and aspires to explanation?

On the face of it, the answer is yes – the renewed role of the popular blockbuster, and an ever growing and insecure emphasis on readership over scholarship, would suggest as much. In Britain, humanist scholars shy away from the metrics that would demonstrate the 'impact' of their work primarily from fear that it may not have any. A single and self-evident instance that evidences a deeper malaise is the failure to cite what we read. We read online journal articles, but cite the hard copy edition; we do keywords searches, while pretending to undertake immersive reading. We search 'Google Books', and pretend we are not.

But even more importantly, we ignore the critical impact of digitisation on our intellectual praxis. [Only 48% of the significant words in the Burney collection of eighteenth-century newspapers are correctly transcribed as a result of poor OCR.\[1\]](#) This makes the other 52% completely un-findable. And of course, from the perspective of the relationship between scholarship and sources, it is always the same 52%. [Bill Turkel](#) describes this as the Las Vegas effect – all bright lights, and an invitation to instant scholarly riches, but with no indication of the odds, and no exit signs. We use the Burney collection regardless – entirely failing to apply the kind of critical approach that historians have built their professional authority upon. This is roulette dressed up as scholarship.

In other words, historians and other humanists have abandoned the rigour of traditional scholarship. Provenance, edition, transcription, editorial practise, readership, authorship, reception – the things academics have traditionally queried in relation to books, are left

unexplored in relation to the online text which now forms the basis of most published history.

As importantly, the way 'history' is promulgated has not kept up either. Why have historians failed to create television programmes with footnotes, and graphs with underlying spreadsheets and sliders? History is part of a grand conversation between the present and the past, played out in extended narrative and analysis, with structure, point, and purpose; but it will be increasingly impoverished if it continues to be produced as a ragged and impotent ghost of a fifteenth century technology. The book had a wonderful 1200 odd year history, which is certainly worth exploring. Its form self-evidently controlled and informed significant aspects of cultural and intellectual change in the West (and through the impositions of Empire, the rest of the world as well); but if historians are to avoid going the way of the book, they need to separate out what they think history is designed to achieve, and to create a scholarly technology that delivers it.

In a rather intemperate attack on the work of Jane Jacobs, published in 1962, Louis Mumford observed that:

... minds unduly fascinated by computers carefully confine themselves to asking only the kind of question that computers can answer and are completely negligent of the human contents or the human results.[2]

In the last couple of decades, historians who are unduly fascinated by books, have restricted themselves to asking only the kind of questions books can answer. Fifty years is a long time in computer science. It is time to find out if a critical and self-consciously scholarly engagement with computers might not now allow the 'human contents' of the past to be more effectively addressed.

A post-endum

This piece was adapted from the rough text of a short talk delivered to a symposium on 'Future Directions in Book History' held at Cambridge University on the 24th of November 2011. It then had an extended afterlife both as a post on my own blog, [Historyonics](#), and in the Open Peer Review section of [Digitalhumanitiesnow.org](#) in preparation for the [Journal of Digital Humanities](#). I then revised it for re-publication in a post-peer review format. The comments were useful, and I am particularly grateful to [John Levin](#), [Adam Crymble](#), [Alycia Sellie](#), [Joe Grobelny](#), and [Lisa Spiro](#) for their willingness to engage critically with it. I have tried to incorporate some of their views within the text. But, I also wanted to take this opportunity to record my own feelings about the process.

The text was originally written in my normal 'ranting' voice, with all the freedom that implies to overstate and shock. The tone is perhaps slightly adolescent, but it is a style that works in the intimate atmosphere of an academic venue, and embeds all the pastiche rhythms and rhetorical ticks I have collected over thirty years of academic writing and lecturing. Its subsequent publication as a blog post was flagged as a text intended for personal, verbal presentation. First person pronouns were retained and the imagined gestures and pauses left to do their work. But in revising it for this post-peer review re-publication I found myself automatically changing it in to a different form, speaking in a different voice – more distant, more careful, more 'academic' for lack of a better word. I have also toned down some (though not all) of the overstatement and hyperbole.

This revision has been an enjoyable process, and I have particularly benefited from the direct engagement with the comments posted, but I am left with yet another conundrum. I like overstatement and hyperbole. I find them intellectually useful, and the form of an un-

reviewed blog and ranty presentation gave me real freedom to indulge in them. The original text reflected all the joys of composing in high voice; and all the freedoms of being an unconstrained publisher of one's own thoughts. In other words, as an author, I was gifted the joy of a blogger, and found that responding to peer review (open or otherwise) merely tarnished and dulled my own pleasure in the product.

Of course, prose is intended for an audience, and preferably an audience that extends beyond the author alone. But this experience makes me wonder if we need to rethink peer review even more fundamentally than the move from closed to open formulae implies. Perhaps we need to recognise that reconstructing a process of selection and revision (of re-creating the scholarly journal online with knobs on) achieves only half the objective. Perhaps we also need to recognise the value of the draft, and the talk; the prose written for an audience of one, and shared only because it can be. Perhaps we need to worry less about the forms and process of generating authority and get on with the work of engaging with a wider world of ideas.

As you will have guessed, I have suddenly moved into blog mode – and it is simply more fun than academic writing.

Originally published by Tim Hitchcock on [October 23, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] Simon Tanner, Trevor Muñoz, and Pich Hemy Ros, "Measuring Mass Text Digitization Quality and Usefulness Lessons Learned from Assessing the OCR Accuracy of the British Library's 19th Century

Online Newspaper Archive," *D-Lib Magazine* 15, no. 7/8 (2009), <http://www.dlib.org/dlib/july09/munoz/o7munoz.html>.

[2] Lewis Mumford, "The Sky Line 'Mother Jacobs Home Remedies'," *The New Yorker* (December 1, 1962), 148.

Defining Data for Humanists: Text, Artifact, Information or Evidence?

Data seems to be the word of the moment for scholarship. The National Endowment for the Humanities and a range of other funders are inviting scholars to “dig data” in their [“Digging into Data” grant program](#). Data itself is now often discussed as representing a [fourth paradigm for scientific discovery and scholarship](#) (PDF). What is a humanist to do in such a situation? Does data, in particular big data, require humanists to adopt a new methodological paradigm? Or, are the kinds of questions humanities scholars traditionally have explored through close reading and hermeneutic interpretation relevant to big data? In this brief essay I suggest some of the ways that humanists already think about and analyze their sources can be employed to understand, explore, and question data.

What is Data to a Humanist?

We can choose to treat data as different kinds of things. First, as constructed things, data are a species of artifact. Second, as authored objects created for particular audiences, data can be interpreted as texts. Third, as computer-processable information, data can be computed in a whole host of ways to generate novel artifacts and texts

which are then open to subsequent interpretation and analysis. Which brings us to evidence. Each of these approaches—data as text, artifact, and processable information—allow one to produce or uncover evidence that can support particular claims and arguments. Data is not in and of itself a kind of evidence but a multifaced object which can be mobilized as evidence in support of an argument.

Data as Constructed Artifacts

Data is always manufactured. It is created. More specifically, data sets are always, at least indirectly, created by people. In this sense, the idea of “raw data” is a bit misleading. The production of a data set requires choices about what and how to collect and how to encode the information. Each of those decisions offers a new potential point of analysis.

Now, when data is transformed into evidence, when we isolate or distill the features of a data set, or when we generate a visualization or present the results of a statistical procedure, we are not presenting the artifact. These are abstractions. The data itself has an [artifactual](#) quality to it. What one researcher considers noise, or something to be discounted in a dataset, may provide essential evidence for another.

In the sciences, there are some tacit and explicit agreements on acceptable assumptions and a set of statistical tests exist to help ensure the validity of interpretations. These kinds of statistical instruments are also great tools for humanists to use. They are not, however, the only way to look at data. For example, the most common use of statistics is to study a small sample in order to make generalizations about a larger population. But statistical tests intended to identify whether trends in small samples scale into larger populations are not

useful if you want to explore the gritty details and peculiarities of a data set.

Data as Interpretable Texts

As a species of human-made artifact, we can think of data sets as having the same characteristics as texts. Data is created for an audience. Humanists can, and should interpret data as an authored work and the intentions of the author are worth consideration and exploration. At the same time, the audience of data also is relevant. Employing a reader-response theory approach to data would require attention to how a given set of data is actually used, understood, and interpreted by various audiences. That could well include audiences of other scientists, the general public, government officials, etc. When we consider what a data set means to individuals within a certain context, we open up a range of fruitful interpretive questions which the humanities are particularly well situated to explicate.

Data as Processable Information

Data can be processed by computers. We can visualize it. We can manipulate it. We can pivot and change our perspective on it. Doing so can help us see things differently. You can process data in a stats package [like R](#) and run a range of statistical tests to uncover statistically significant differences or surface patterns and relationships. Alternatively, you can deform a data set with a process like [Spoonbill's N+7 machine](#), which replaces every noun in a text with the seventh word in the dictionary that follows the original, thus prompting you to see the original data from a different perspective, as Mark Sample's [Hacking the Accident](#) did for [Hacking the Academy](#). In both cases, you can process information—numerical or textual—to change your frame of understanding for a particular set of data.

Importantly, the results of processed information are not necessarily declarative answers for humanists. If we take seriously Stephen Ramsay's suggestions for [algorithmic criticism](#), then data offers humanists the opportunity to manipulate or algorithmically derive or generate new artifacts, objects, and texts that we also can read and explore.^[1] For humanists, the results of information processing are open to the same kinds of hermeneutic exploration and interpretation as the original data.

Data Can Hold Evidentiary Value

As a species of human artifact, as a cultural object, as a kind of text, and as processable information, data is open to a range of hermeneutic tactics for interpretation. In much the same way that encoding a text is an interpretive act, so are creating, manipulating, transferring, exploring, and otherwise making use of data sets. Therefore, data is an artifact or a text that can hold the same potential evidentiary value as any other kind of artifact. That is, scholars can uncover information, facts, figures, perspectives, meanings, and traces of thoughts and ideas through the analysis, interpretation, exploration, and engagement with data, which in turn can be deployed as evidence to support all manner of claims and arguments. I contend that data is not a kind of evidence; it is a potential source of information that can hold evidentiary value.

Conclusion

Approaching data in this way should feel liberating to humanists. For us, data and the capabilities of processing data are not so much new methodological paradigms, rather an opportunity for us to bring the skills we have honed in the close reading of texts and artifacts into service for this new species of text and artifact. Literary scholar Franco Moretti already has asked us to pivot, to begin to engage in [distant](#)

[reading](#). What should reassure us all is that at the end of the day, any attempt at distant reading results in a new artifact that we can also read closely.

In the end, the kinds of questions humanists ask about texts and artifacts are just as relevant to ask of data. While the new and exciting prospects of processing data offer humanists a range of exciting possibilities for research, humanistic approaches to the textual and artifactual qualities of data also have a considerable amount to offer to the interpretation of data.

Originally published by Trevor Owens on [December 15, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Champaign: University of Illinois Press, 2011).

Demystifying Networks, Parts I & II

Part 1 of n : An Introduction

This piece builds on a [bunch of my recent blog posts](#) that have mentioned networks. [Elijah Meeks](#) already has prepared a good introduction to [network visualizations on his own blog](#), so I cover more of the conceptual issues here, hoping to reach people with little-to-no background in networks or math, and specifically to digital humanists interested in applying network analysis to their own work.

Some Warnings

A network is a fantastic tool in the digital humanist's toolbox—one of many—and it's no exaggeration to say pretty much *any* data can be studied via network analysis. With enough stretching and molding, you too could have a network analysis problem! As with many other science-derived methodologies, it's fairly easy to extend the metaphor of network analysis into any number of domains.

The danger here is two-fold.

1. [When you're given your first hammer, everything looks like a nail.](#)
Networks *can* be used on any project. Networks *should* be used on

far fewer. Networks in the humanities are experiencing quite the awakening, and this is due in part to the until-recently untapped resources of easy tools and available datasets. There is a lot of low-hanging fruit out there on the networks+humanities tree, and they ought to be plucked by those brave and willing enough to do so. However, that does not give us an excuse to apply networks to *everything*. This series will talk a little bit about when hammers are useful, and when you really should be reaching for a screwdriver.

2. Methodology appropriation is *dangerous*. Even when the people designing a methodology for some specific purpose get it right—and they rarely do—there is often a score of theoretical and philosophical caveats that get lost when the methodology gets translated. In the more frequent case, when those caveats are not known to begin with, “borrowing” the methodology becomes even more dangerous. Ted Underwood [blogs a great example of why literary historians ought to skip a major step in Latent Semantic Analysis](#), because the purpose of the literary historian is *so very different* from that of the computer scientist who designed the algorithm. This series will attempt to point out some of the theoretical baggage and necessary assumptions of the various network methods it covers.

The Basics

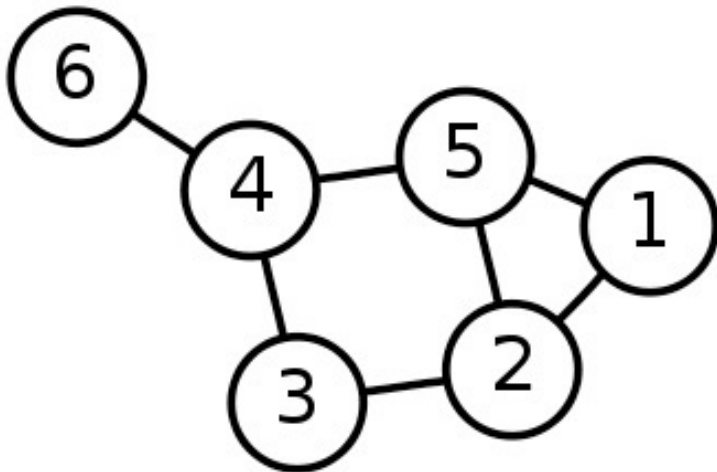
Nothing worth discovering has ever been found in safe waters. Or rather, everything worth discovering in safe waters *has already been discovered*, so it's time to shove off into the dangerous waters of methodology appropriation, cognizant of the warnings but not crippled by them.

Anyone with a lot of time and a vicious interest in networks should stop reading *right now*, and instead pick up copies of [Networks, Crowds, and Markets](#)[1] and [Networks: An Introduction](#)[2]. The first is a non-mathy introduction to most of the concepts of network analysis, and the second is a more in-depth (and formula-laden) exploration of those concepts. They're phenomenal, essential, and worth every penny.

Those of you with slightly less time, but somehow enough to read my rambling blog (there are apparently a few of you out there), so good of you to join me. We'll start with the *really basic* basics, but stay with me, because by part *n* of this series, we'll be going over the really cool stuff only ninjas, Gandhi, and The Rolling Stones have worked on.

Networks

The word “network” originally meant just that: “a [net-like arrangement of threads, wires, etc.](#)” It later came to stand for any complex, interlocking system. **Stuff** and **relationships**.



A simple-network representation from Wikipedia.org

Generally, network studies are made under the assumption that neither the stuff nor the relationships are the whole story on their own. If you're studying something with networks, odds are you're doing so because you think the objects of your study are *interdependent* rather than *independent*. Representing information as a network implicitly suggests not only that connections matter, but that they are *required* to understand whatever's going on.

Oh, I should mention that people often use the word “graph” when talking about networks. It's basically the mathy term for a network, and its definition is a bit more formalized and concrete. Think dots connected with lines.

Because networks are studied by lots of different groups, there are lots of different words for pretty much the same concepts. I'll explain some of them below.

The Stuff

Stuff (presumably) exists. Eggplants, true love, the *Mary Celeste*, tall people, and Terry Pratchett's *Thief of Time* all fall in that category. Network analysis generally deals with one or a small handful of *types* of stuff, and then a multitude of examples of that type.

Say the *type* we're dealing with is a book. While scholars might argue the exact lines of demarcation separating book from non-book, I think we can all agree that most of the stuff on my bookshelf are, in fact, books. They're the *stuff*. There are different examples of books: a quotation dictionary, a Poe collection, and so forth.

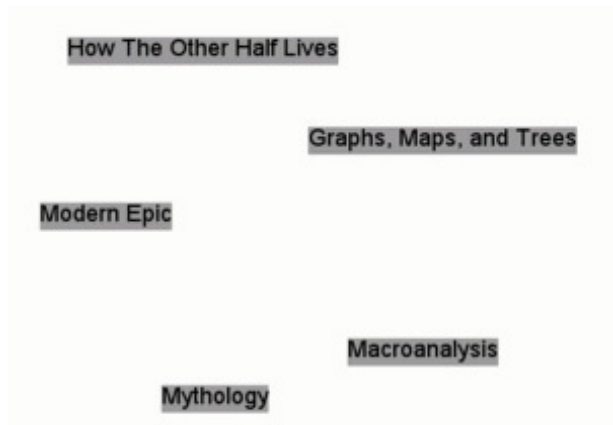
I'll call this assortment of stuff *nodes*. You'll also hear them called *vertices* (mostly from the mathematicians and computer scientists), *actors* (from the sociologists), *agents* (from the modelers), or *points* (not really sure where this one comes from).

The *type* of stuff corresponds to the *type* of node. The individual examples are the nodes themselves. All of the nodes are books, and each book is a different node.

Nodes can have attributes. Each node, for example, may include the title, the number of pages, and the year of publication.

A list of nodes could look like this:

Title	# of pages	year of publication
Graphs, Maps, and Trees	119	2005
How The Other Half Lives	233	1890
Modern Epic	272	1995
Mythology	352	1942
Macroanalysis	unknown	2011



A network of books (nodes) with no relationships (connections)

We can get a bit more complicated and add more node *types* to the network. Authors, for example. Now we've got a network with books

and authors (but nothing linking them, yet!). *Franco Moretti* and *Graphs, Maps, and Trees* are both nodes, although they are of different varieties, and not yet connected. We could have a second list of nodes, part of the same network, that might look like this:

Author	Birth	Death

Franco Moretti	?	n/a
Jacob A. Riis	1849	1914
Edith Hamilton	1867	1963
Matthew Jockers	?	n/a



A network of books and authors without relationships

A network with two types of nodes is called *2-mode*, *bimodal*, or *bipartite*. We can add more, making it *multimodal*. Publishers, topics, you-name-it. We can even add seemingly unrelated node-types, like academic conferences, or colors of the rainbow. The list goes on. We would have a new list for each new variety of node.

Presumably we could continue adding nodes and node-types until we run out of stuff in the universe. This would be a bad idea, and not just because it would take more time, energy, and hard-drives than could ever possibly exist. As it stands now, network science is ill-equipped to deal with multimodal networks. 2-mode networks are difficult enough to work with, but once you get to three or more varieties of nodes, most algorithms used in network analysis *simply do not work*. It's not that they *can't* work; it's just that most algorithms were only created to deal with networks with one variety of node. This is a trap I see many newcomers to network science falling into, especially in the digital humanities. They find themselves with a network dataset of, for example, authors and publishers. Each author is connected with one or several publishers (we'll get into the connections themselves in the next section), and the up-and-coming network scientist loads the network into their favorite software and visualizes it. Woah! A network! Then, because the software is easy to use, and has a lot of buttons with words that from a non-technical standpoint seem to make a lot of sense, they press those buttons to see what comes out. Then, they change the visual characteristics of the network based on the buttons they've pressed. Let's take a concrete example. Popular network software [Gephi](#) comes with a button that measures the *centrality* of nodes. Centrality is a pretty complicated concept that I'll get into more detail later, but for now it's enough to say that it does exactly what it sounds like: it finds how central, or important, each node is in a network. The newcomer to network analysis loads the author-publisher network into Gephi, finds the centrality of every

node, and then makes the nodes bigger that have the highest centrality. The issue here is that, although the network loads into Gephi perfectly fine, and although the centrality algorithm runs smoothly, the resulting numbers *do not mean what they usually mean*. Centrality, as it exists in Gephi, was fine-tuned to be used with single mode networks, whereas the author-publisher network (not to mention the author-book network above) is bimodal. Centrality measures have been made for bimodal networks, but those algorithms are not included with Gephi. Most computer scientists working with networks do so with only one or a few types of nodes. Humanities scholars, on the other hand, are often dealing with the interactions of *many* types of things, and so the algorithms developed for traditional network studies are insufficient for the networks we often have. There are ways of fitting their algorithms to our networks, or vice-versa, but that requires fairly robust technical knowledge of the task at hand. Besides dealing with the single mode / multimodal issue, humanists also must struggle with fitting square pegs in round holes. Humanistic data are almost by definition uncertain, open to interpretation, flexible, and not easily definable. Node types are by definition concrete; your object either *is* or *is not* a book. Every book-type thing must share certain unchanging characteristics. This *reduction of data* comes at a price, one that some argue traditionally divided the humanities and social sciences. If humanists care more about the differences than the regularities, more about what makes an object unique rather than what makes it similar, that is the very information they are likely to lose by defining their objects as nodes. This is not to say it cannot be done, or even that it has not! People are clever, and network science is more flexible than some give it credit for. The important thing is either to be aware of what you are losing when you reduce your objects to one or a few types of nodes, or to change the methods of network science to fit your more complex data.

The Relationships

Relationships (presumably) exist. Friendships, similarities, web links, authorships, and wires all fall into this category. Network analysis generally deals with one or a small handful of *types* of relationships, and then a multitude of examples of that type. Now that we have *stuff* and *relationships*, we're equipped to represent everything needed for a simple network. Let's start with a single mode network; that is, a network with only one sort of node: cities. We can create a network of which cities are connected to one another by at least one single stretch of highway, like the one below:

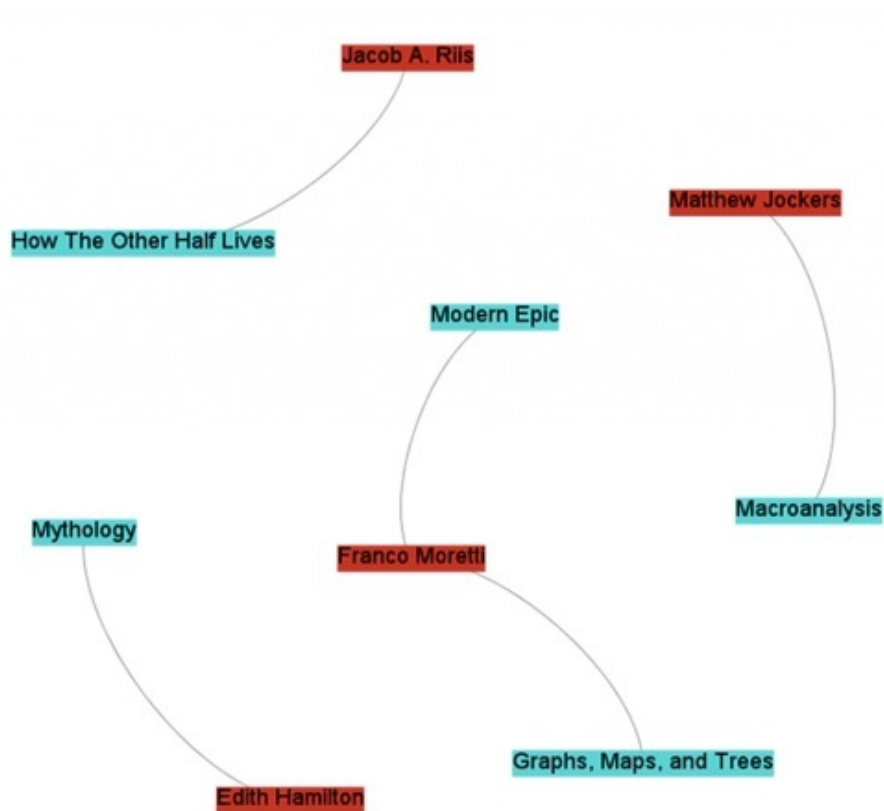
City	is connected to
Indianapolis	Louisville
Louisville	Cincinnati
Cincinnati	Indianapolis
Cincinnati	Lexington
Louisville	Lexington
Louisville	Nashville



Cities interconnected by highways

The simple network above shows how certain cities are connected to one another via highways. A connection via a highway is the *type* of relationship. An example of one of the above relationships can be stated "Louisville *is connected via a highway to* Indianapolis." These connections are *symmetric* because a connection from Louisville to Indianapolis also implies a connection in the reverse direction, from Indianapolis to Louisville. More on that shortly. First, let's go back to the example of books and authors from the last section. Say the *type* we're dealing with is an authorship. Books (the *stuff*) and authors (another kind of *stuff*) are connected to one-another via the authorship relationship, which is formalized in the phrase "X is an author of Y." The individual relationships themselves are of the form "Franco Moretti is an author of *Graphs, Maps, and Trees*." Much like the *stuff* (nodes), relationships enjoy a multitude of names. I'll call them *edges*. You'll also hear them called *arcs*, *links*, *ties*, and *relations*. For simplicity sake, although *edges* are often used to describe only one variety of relationship, I'll use it for pretty much everything and just add qualifiers when discussing specific types. The *type* of relationship corresponds to the *type* of edge. The individual examples are the edges themselves. Individual edges are defined, in part, by the nodes that they connect. A list of edges could look like this:

Person	Is an author of
Franco Moretti	Modern Epic
Franco Moretti	Graphs, Maps, and Trees
Jacob A. Riis	How The Other Half Lives
Edith Hamilton	Mythology
Matthew Jockers	Macroanalysis



Network of books, authors, and relationships between them

Notice how, in this scheme, edges can only link two different types of nodes. That is, a person can be an author of a book, but a book cannot be an author of a book, nor can a person an author of a person. For a network to be truly bimodal, it *must* be of this form. Edges can go between types, but not among them. This constraint may seem artificial, and in some sense it is, but for now the short explanation is that it is a constraint required by most algorithms that deal with bimodal networks. As mentioned above, algorithms are developed for specific purposes. Single mode networks are the ones with the most research done on them, but bimodal networks certainly come in a close

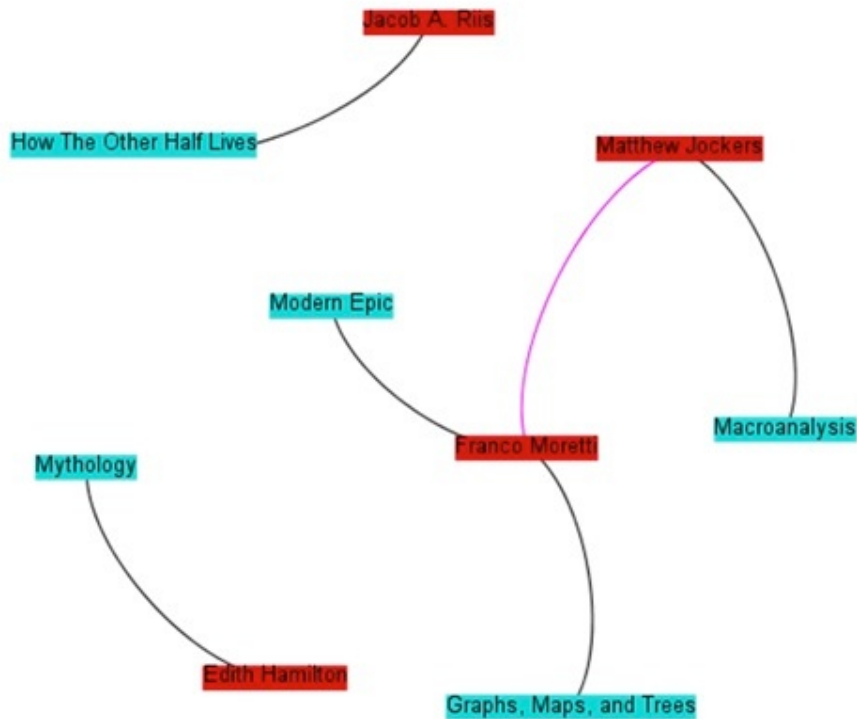
second. They are networks with two types of nodes, and edges *only* going *between* those types. Contrast this against the single mode city-to-city network from before, where edges connected nodes of the same type. Of course, the world humanists care to model is often a good deal more complicated than that, and not only does it have multiple varieties of nodes – it also has multiple varieties of edges. Perhaps, in addition to “X is an author of Y” type relationships, we also want to include “A collaborates with B” type relationships. Because edges, like nodes, can have attributes, an edge list combining both might look like this.

Node1	Node 2	Edge Type
-----	-----	-----
Franco Moretti	Modern Epic	is an author of
Franco Moretti	Graphs, Maps, and Trees	is an author of
Jacob A. Riis	How The Other Half Lives	is an author of
Edith Hamilton	Mythology	is an author of
Matthew Jockers	Macroanalysis	is an author of
Matthew Jockers	Franco Moretti	collaborates with

Notice that there are now two types of edges: “is an author of” and “collaborates with.” Not only are they two different types of edges; they act in two *fundamentally different ways*. “X is an author of Y” is an asymmetric relationship; that is, you cannot switch out Node1 for Node2. You cannot say “Modern Epic is an author of Franco Moretti.” We call this type of relationship a *directed edge*, and we generally represent that visually using an arrow going from one node to another.

“A collaborates with B,” on the other hand, is a symmetric relationship. We can switch out “Matthew Jockers collaborates with Franco Moretti” with “Franco Moretti collaborates with Matthew Jockers,” and the information represented would be exactly the same. This is called an *undirected edge*, and is usually represented visually by a simple line

connecting two nodes. Notice that this is an edge connecting two nodes of the same type (an author-to-author connection), and recall that true bimodal networks require edges to only go *between* types. Algorithms meant for bimodal networks no longer apply to the network above.



Network of authors, books, authorship relationships, and collaboration relationships

Most network algorithms and visualizations break down when combining these two flavors of edges. Some algorithms were designed for directed edges, like [Google's PageRank](#), whereas other algorithms are designed for undirected edges, like many centrality measures. Combining both types is rarely a good idea. Some algorithms will still

run when the two are combined, however the results usually make little sense.

Both directed and undirected edges can also be weighted. For example, I can try to make a network of books, with those books that are similar to one another sharing an edge between them. The more similar they are, the heavier the weight of that edge. I can say that every book is similar to every other on a scale from 1 to 100, and compare them by whether they use the same words. Two dictionaries would probably connect to one another with an edge weight of 95 or so, whereas *Graphs, Maps, and Trees* would probably share an edge of weight 5 with *How The Other Half Lives*. This is often visually represented by the thickness of the line connecting two nodes, although sometimes it is represented as color or length.

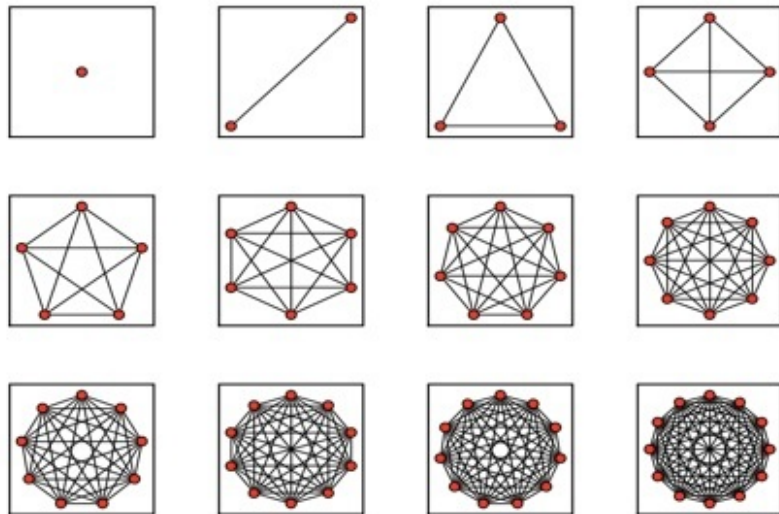
It's also worth pointing out the difference between explicit and inferred edges. If we're talking about computers connected on a network via wires, the edges connecting each computer *actually exist*. We can weight them by wire length, and that length, too, *actually exists*. Similarly, citation linkages, neighbor relationships, and phone calls are explicit edges.

We can begin to move into interpretation when we begin creating edges between books based on similarity (even when using something like word comparisons). The edges are a layer of interpretation not intrinsic in the objects themselves. The humanist might argue that all edges are intrinsic all the way down, or inferred all the way up, but in either case there is a difference in kind between two computers connected via wires, and two books connected because we feel they share similar topics.

As such, algorithms made to work on one may not work on the other; or perhaps they may, but their interpretative framework must change

drastically. A very central computer might be one in which, if removed, the computers will no longer be able to interact with one another; a very central book may be something else entirely.

As with nodes, edges come with many theoretical shortcomings for the humanist. Really, everything is probably related to everything else in its [light cone](#). If we've managed to make everything in the world a node, realistically we'd also have some sort of edge between pretty much everything, with a lesser or greater weight. A network of nodes where almost everything is connected to almost everything else is called *dense*, and dense networks are rarely useful. Most network algorithms (especially ones that detect communities of nodes) work better and faster when the network is *sparse*, when most nodes are only connected to a small percentage of other nodes.



Maximally dense networks from sagemath.org

To make our network sparse, we often must artificially cut off which edges to use, especially with humanistic and inferred data. That's what [Shawn Graham showed us how to do when combining topic models with networks](#). The network was one of authors and topics; which authors wrote about which topics? The data itself connected every author to every topic to a greater or lesser degree, but such a dense network would not be very useful, so Shawn limited the edges to the *highest weighted* connections between an author and a topic. The resulting network looked like [this](#) (PDF), when it otherwise would have looked like a big ball of spaghetti and meatballs.

Unfortunately, given that humanistic data are often uncertain and biased to begin with, every arbitrary act of data-cutting has the potential to add further uncertainty and bias to a point where the network no longer provides meaningful results. The ability to cut away just enough data to make the network manageable, but not enough to lose information, is as much an art as it is a science.

Hypergraphs & Multigraphs

Mathematicians and computer scientists have actually formalized more complex varieties of networks, and they call them [hypergraphs](#) and [multigraphs](#). Because humanities data are often so rich and complex, it may be more appropriate to represent them using these representations. Unfortunately, although ample research has been done on both, most out-of-the-box tools support neither. We have to build them for ourselves.

A hypergraph is one in which more than two nodes can be connected by one edge. A simple example would be an "is a sibling of" relationship, where the edge connected three sisters rather than two. This is a symmetric, undirected edge, but perhaps there can be directed

edges as well, of the type “Alex *convinced* Betty to run away from Carl.” A three-part edge.

A multigraph is one in which multiple edges can connect any two nodes. We can have, for example, a transportation graph between cities. A edge exists for every transportation route. Realistically, many routes can exist between any two cities: some by plane, several different highways, trains, etc.

I imagine both of these representations will be important for humanists going forward, but rather than relying on that computer scientist who keeps hanging out in the history department, we ourselves will have to develop algorithms that accurately capture exactly what it is we are looking for. We have a different set of problems, and though the solutions may be similar, they must be adapted to our needs.

Side note: RDF Triples

Digital humanities loves [RDF](#) (Resource Description Framework), which is essentially a method of storing and embedding structured data. RDF basically works using something called a *triple*; a subject, a predicate, and an object. “Moretti is an author of *Graphs, Maps, and Trees*” is an example of a triple, where “Moretti” is the subject, “is an author of” is the predicate, and “*Graphs, Maps, and Trees*” is the object. As such, nearly all RDF documents can be represented as a directed network. Whether that representation would actually be useful depends on the situation.

Side note: Perspectives

Context is key, especially in the humanities. One thing the last few decades has taught us is that perspectives are essential, and any model

of humanity that does not take into account its multifaceted nature is doomed to be forever incomplete. According to Alex, his friends Betty and Carl are best friends. According to Carl, he can’t actually stand Betty. The structure and nature of a network might change depending on the perspective of a particular node, and I know of no model that captures this complexity. If you’re familiar with something that might capture this, or are working on it yourself, please let me know via e-mail.

Networks, Revisited

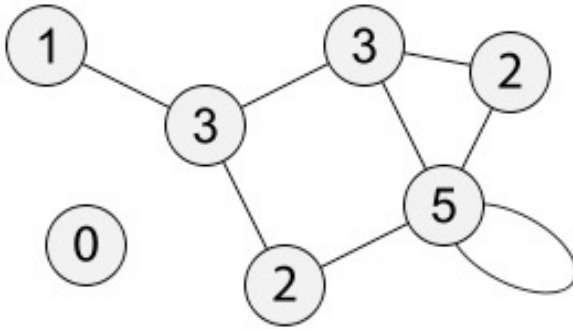
This piece has discussed the simplest units of networks: the stuff and the relationships that connect them. Any network analysis approach must subscribe to and live with that duality of objects. Humanists face problems from the outset: data that do not fit neatly into one category or the other, complex situations that ought not be reduced, and methods that were developed with different purposes in mind. However, network analysis remains a viable methodology for answering and raising humanistic questions—we simply must be cautious, and must be willing to get our hands dirty editing the algorithms to suit our needs.

Part II: Node Degree: An Introduction

In Part II, I will cover the deceptively simple concept of *node degree*. I say “deceptive” because, on the one hand, network degree can tell you quite a lot. On the other hand, degree can often lead one astray, especially as networks become larger and more complicated.

A node’s *degree* is, simply, how many edges it is connected to. Generally, this also correlates to how many *neighbors* a node has, where a node’s neighborhood is those other nodes connected directly

to it by an edge. In the network below, each node is labeled by its degree.



Each node in the network is labeled with its degree, from Wikipedia.org

If you take a minute to study the network, something might strike you as odd. The bottom-right node, with degree 5, is connected to only four distinct edges, and really only three other nodes (four, including itself). *Self-loops*, which will be discussed later, are counted twice. A self-loop is any edge which starts and ends at the same node.

Why are self-loops counted twice? Well, as a rule of thumb you can say that, since the degree is the number of times the node is connected to an edge, and a self-loop connects to a node twice, that's the reason. There are some more math-y reasons dealing with matrix representation, another topic for a later date. Suffice it to say that many network algorithms will not work well if self-loops are only counted once.

The odd node out on the bottom left, with degree zero, is called an *isolate*. An isolate is any node with no edges.

At any rate, the concept is clearly simple enough. Count the number of times a node is connected to an edge, get the degree. If only getting higher education degrees were this easy.

Centrality

Node degree is occasionally called *degree centrality*. [Centrality](#) is generally used to determine how important nodes are in a network, and lots of clever researchers have come up with lots of clever ways to measure it. "Importance" can mean a lot of things. In social networks, centrality can be the amount of influence or power someone has; in the U.S. electrical grid network, centrality might mean which power station should be removed to cause the most damage to the network.

The simplest way of measuring node importance is to just look at its degree. This centrality measurement at once seems deeply intuitive and extremely silly. If we're looking at the social network of [Facebook](#), with every person a node connected by an edge to their friends, it's no surprise that the most well-connected person is probably also the most powerful and influential in the social space. On the same token, though, degree centrality is such a coarse-grained measurement that it's really anybody's guess what *exactly* it's measuring. It could mean someone has a lot of power; it could also mean that someone tried to become friends with absolutely everybody on Facebook. Recall the example of a city-to-city network from Part I of this series: Louisville was the most central city because you have to drive through it to get to the most others.

Degree Centrality Sampling Warnings

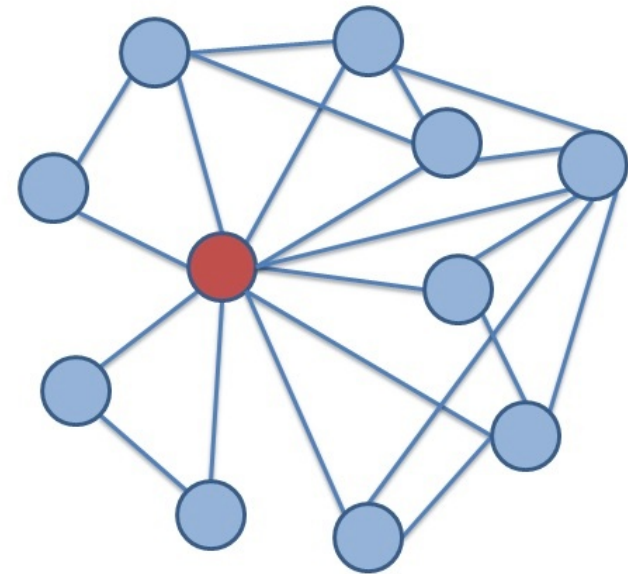
Degree works best as a measure of network centrality when you have *full knowledge* of the network. That is, a social network exists, and instead of getting some glimpse of it and analyzing just that, you

have the entire context of the social network: all the friends, all the friends of friends, and so forth.

When you have an *ego-network* (a network of one person, like a list of all my friends and who among them are friends with one another), clearly the node with the highest centrality is the ego node itself. This knowledge tells you very little about whether that ego is actually central within the larger network, because you sampled the network *such that the ego is necessarily the most central*. Sampling strategies—how you pick which nodes and edges to collect—can fundamentally affect centrality scores. The city-to-city network from Part I has Louisville as the most central city, however a simple look at a map of the United States would show that, given more data, this would no longer be the case.

A historian of science might generate a correspondence network from early modern letters currently held in Oxford's library. In fact, this is currently happening, and the resulting resource will be invaluable. Unfortunately, centrality scores generated from nodes in that early modern letter writing network will more accurately reflect the whims of Oxford editors and collectors over the years, rather than the underlying correspondence network itself. Oxford scholars over the years selected certain collections of letters, be they from Great People or sent to or from Oxford, and that choice of what to hold at Oxford libraries will bias centrality scores toward Oxford-based scholars, Great People, and whatever else was selected for.

Similarly, the generation of a social network from a literary work will bias the recurring characters; characters that occur more frequently are simply statistically more likely to appear with more people, and as such will have the highest degrees. It is likely that the degree centrality and frequency of character occurrence are almost exactly correlated.



An ego network from Wikipedia.org

Of course, if what you're looking for is *the most central character in the novel* or *the most central figure from Oxford's perspective*, this measurement might be perfectly sufficient. The important thing is to be aware of the limitations of degree centrality, and the possible biasing effects from selection and sampling. Once those biases are explicit, careful and useful inferences can still be drawn.

Things get a bit more complicated when looking at document similarity networks. If you've got a network of books with edges connecting them based on whether they share similar topics or keywords, your degree centrality score will mean something *very different*. In this case, centrality could mean the most general book. Keep in mind that book length might affect these measurements as well; the longer a book is, the more likely (by chance alone) it will cover more topics. Thus, longer

books may also appear to be more central, if one is not careful in generating the network.

Degree Centrality in Bimodal Networks

Recall that bimodal networks are ones where there are two different types of nodes (e.g., articles and authors), and edges are relationships that bridge those types (e.g., authorships). In this example, the more articles an author has published, the more central she is. Degree centrality would have nothing to do, in this case, with the number of co-authorships, the position in the social network, etc.

With an even more multimodal network, having many types of nodes, degree centrality becomes even less well defined. As the sorts of things a node can connect to increases, the utility of simply counting the number of connections a node has decreases.

Micro vs. Macro

Looking at the degree of an individual node, and comparing it against others in the network, is useful for finding out about the relative position of that node within the network. Looking at the degree of *every node at once* turns out to be exceptionally useful for talking about the network as a whole, and comparing it to others. I'll leave a thorough discussion of degree distributions for a later post, but it's worth mentioning them in brief here. The degree distribution shows how many nodes have how many edges.

As it happens, many real world networks exhibit something called "power-law properties" in their degree distributions. What this essentially means is that a small number of nodes have an exceptionally high degree, whereas most nodes have very low degrees. By comparing the degree distributions of two networks, it is possible to

say whether they are structurally similar. There's been some fantastic work comparing the degree distribution of social networks in various plays and novels to find if they are written or structured similarly.

Extending Degree

For the entirety of this piece, I have been talking about networks that were unweighted and undirected. Every edge counted just as much as every other, and they were all *symmetric* (a connection from A to B implies the same connection from B to A). Degree can be extended to both weighted and directed (*asymmetric*) networks with relative ease.

Combining degree with edge weights is often called *strength*. The strength of a node is the sum of the weights of its edges. For example, let's say Steve is part of a weighted social network. The first time he interacts with someone, an edge is created to connect the two with a weight of 1. Every subsequent interaction incrementally increases the weight by 1, so if he's interacted with Sally four times, Samantha two times, and Salvador six times, the edge weights between them are 4, 2, and 6 respectively.

In the above example, because Steve is connected to three people, his degree is $1+1+1=3$. Because he is connected to one of them four times, another twice, and another six times, his weight is $4+2+6=8$.

Combining degree with directed edges is also quite simple. Instead of one degree score, every node now has two different degrees: *in-degree* and *out-degree*. The in-degree is the number of edges pointing to a node, and the out-degree is the number of edges pointing away from it. If Steve *borrowed* money from Sally, and *lent* money to Samantha and Salvador, his in-degree might be **1** and his out-degree **2**.

Powerful Degrees

The degree of a node is really very simple: more connections, higher degree. However, this simple metric accounts for quite a great deal in network science. Many algorithms that analyze both node-level properties and network-level properties are closely correlated with degree and degree distribution. This is a [pareto](#)-like effect; a great deal about a network is driven by the degree of its nodes.

While degree-based results are often intuitive, it is worth pointing out that the prime importance of degree is a direct result of the binary network representation of nodes and edges. Interactions either happen or they don't, and everything that *is* is a self-contained node or edge. Thus, how many nodes, how many edges, and which nodes have which edges will be the driving force of any network analysis. This is both a limitation and a strength; basic counts influence so much, yet they are apparently powerful enough to yield intuitive, interesting, and ultimately useful results.

Originally published by Scott Weingart on [December 14, 2011](#) and [December 17, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

I plan to continue blogging about network analysis, so if you have any requests, please feel free to get in touch with me at scbweing at indiana dot edu.

Notes:

[1] David Easley and Jon M. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World* (Cambridge: Cambridge University Press, 2010).

[2] Mark E. J. Newman, *Networks: An Introduction*, 1st ed (Oxford: Oxford University Press, 2010).

Clustering with Compression for the Historian

INTRODUCTION

I mentioned [in my blog](#) that I'm playing around with a variety of clustering techniques to identify patterns in legal records from the early modern Spanish Empire. In this post, I will discuss the first of my training experiments using Normalized Compression Distance (NCD). I'll look at what NCD is, some potential problems with the method, and then the results from using NCD to analyze the Criminales Series descriptions of the Archivo Nacional del Ecuador's ([ANE](#)) Series Guide. For what it's worth, this is a very easy and approachable method for measuring similarity between documents and requires almost no programming chops. So, it's perfect for me!

WHAT IS NCD?

I was inspired to look at NCD for clustering by a pair of posts by Bill Turkel ([here](#), [here](#)) from quite a few years ago. Bill and Stéfan Sinclair also [used NCD](#) to cluster cases for the [Digging Into Data Old Bailey Project](#). Turkel's posts provide a nice overview of the method, which was proposed in 2005 by Rudi Cilibrasi and Paul Vitányi.^[1] Essentially,

Cilibrasi and Vitányi proposed measuring the distance between two strings of arbitrary length by comparing the sum of the lengths of the individually compressed files to a compressed concatenation of the two files. So, adding the compressed length of x to the compressed length of y will be longer than the compressed length of $(x|y)$. How much longer is what is important. The formula is this, where $c(x)$ is the length of x compressed:

$$\text{NCD}(x, y) = [C(x|y) - \min\{C(x), C(y)\}] / \max\{C(x), C(y)\}$$

$C(x|y)$ is the compression of the concatenated strings. Theoretically, if you concatenated and compressed two identical strings, you would get a distance of 0 because $[C(x|x) - C(x)]/C(x)$ would equal 0/1, or 0. As we'll see in a bit, though, this isn't the case and the overhead required by the various compression algorithms at our disposal make a 0 impossible, and more so for long strings depending on the method. Cilibrasi and Vitányi note that in practice, that if r is the NCD, the NCD will be $0 \leq r \leq 1 + \epsilon$, where ϵ is usually around 0.1, and accounts for the implementation details of the compression algorithm. Suffice to say, though, that the closer to 0 the result is, the more similar the strings (or files in our case) are. Nonetheless, the distance between two strings, or files, or objects as measured with this formula can then be used to cluster those strings, files, or objects. One obvious advantage to the method is that it works for comparing strings of arbitrary length with one another.

Why does this work? Essentially, lossless compression suppresses redundancy in a string, while maintaining the ability to fully restore the file. Compression algorithms evolved to deal with constraints in the storage and transmission of data. It's easy to forget in the age of the inexpensive terabyte hard drive what persistent storage once cost. In 1994, the year that the first edition of Witten, Moffat, and

Bell's [Managing Gigabytes](#) was published, hard disk storage still ran at close to [\\$1/megabyte](#). That's right, just 17 years ago that 500GB drive in your laptop would have cost \$500,000. To put that into perspective, in [1980 IBM](#) produced one of the first disk drives to break the GB barrier. The [2.52GB IBM 3380](#) was initially released in 5 different models, and ranged in price between \$81,000 and \$142,000. For what it's worth, the median housing price in Washington, DC in 1980 was the second highest in the country at \$62,000. A hard disk that cost twice as much as the median house in DC. Obviously not a consumer product. At the per/GB rate that the 3380 sold for, your 500GB drive would have cost up to \$28,174,603.17! In inflation-adjusted dollars for 2011 that would be \$80.5M! An absurd comparison, to be sure. Given those constraints, efficiency in data compression made real dollars sense. Even still, despite the plunging costs of storage and growing bandwidth capacity, text and image compression remains an imperative in computer science.

As Witten, et al. define it,

Text compression ... involves changing the representation of a file so that it takes less space to store or less time to transmit, yet the original file can be reconstructed exactly from the compressed representation.[2]

This is lossless compression (as opposed to lossy compression, which you may know from messing with jpegs or other image formats). There are a variety of compression methods, each of which takes a different approach to compressing text data and which are either individually or in some kind of combination behind the compression formats you're used to—.zip, .bz2, .rar, .gz, etc. Frequently, they also have their roots in the early days of electronic data. [Huffman coding](#) was developed by an eponymous MIT graduate student in the early 1950s.

In any case, the objective of a compression method is to locate, remove, store, and recover redundancies within a text. NCD works because within a particular algorithm, the compression method is consistently imposed on the data, thus making the output comparable. What isn't comparable, though, is mixing algorithms.

LIMITATIONS: SIZE MATTERS

Without getting too technical (mostly because I get lost once it goes too far), it's worth noting some limitations based on which method of compression you chose when applying NCD. Shortly after Cilibrasi and Vitányi published their paper on clustering via compression, Cebrián, et al. published a piece that compared the integrity of NCD between three compressors— bzip2, gzip, and PPMZ.[3] The paper is interesting, in part, because the authors do an excellent job of explaining the mechanics of the various compressors in language that even I could understand.

I came across this paper through some google-fu because I was confused by the initial results I was getting while playing around with my Criminales Series Guide. Python has built-in support for compression and decompression using bzip2 and gzip, so that's what I was using. I have the Criminales Series divided into decades from 1601 to 1830. My script was walking through and comparing every file in the directory to every other one, including itself. I assumed that the concatenation of two files that were identical would produce a distance measurement of 0, and was surprised to see that it wasn't happening, and in some cases not even close. (I also hadn't read much of anything about compression at that point!) But that wasn't the most surprising thing. What was more surprising was that in the latter decades of my corpus, the distance measures when comparing individual decades to themselves were actually coming out very high. Or, at least they were

using the gzip algorithm. For example, the decade with the largest number of cases, and thus the longest text, is 1781-1790 at about 39,000 words. Gzip returned an NCD of 0.97458 when comparing this decade to itself. What? How is that possible?

Cebrián, et al. explain how different compression methods have upper limits to the size of a block of text that they operate on before needing to break that block into new blocks. This makes little difference from the perspective of compressors doing their job, but it does have implications for clustering. The article goes into more detail, but here's a quick and dirty overview.

bzip2

The bzip2 compressor works in three stages to compress a string: (1) a [Burrows-Wheeler Transform](#), (2) a [move-to-front transform](#), and (3) a statistical compressor like Huffman coding.^[4] The bzip2 algorithm can perform this method on blocks of text up 900KB without needing to break the block of text into two blocks. So, for NCD purposes, this means that if a pair of files are concatenated, and the size of this pair is less than 900KB, what the bzip compressor will see is essentially a mirrored text. But, if the concatenated file is larger than 900KB, then bzip will break the concatenation into more than one block, each of which will be sent through the three stages of compression. But, these blocks will no longer be mirrors. As a result, the NCD will cease to be robust. Cebrián, et al. claim that the NCD for $C(x|x)$ should fall in a range between **0.2 and 0.3**, and anything beyond that indicates it's not a good choice for comparing the set of documents under evaluation.

gzip

The gzip compressor uses a different method than bzip2's block compression, one based on the Lempel-Ziv [LZ77](#) algorithm, also

known as **sliding window compression**. Gzip then takes the LZ77-processed string and subjects it to a statistical encoding like Huffman. It's the first step that is important for us, though. Sliding window compression searches for redundancies by taking 32KB blocks of data, and looking ahead at the next 32KB of data. The method is much faster than bzip2's block method. (In my experiments using python's zlib module, code execution took about 1/2 the time as python's bzip on default settings.) And, if the text is small, such that $C(x|x) < 32\text{KB}$, the NCD result is better. Cebrián, et al. find that gzip returns an NCD result in the range between 0 and 0.1. But, beyond 32KB they find that NCD rapidly grows beyond 0.9 — exactly what I saw with the large 1781-1790 file (which is 231KB).

lzma

Cebrián, et al. offer a third compressor, ppmz, as an alternative to bzip2 and gzip for files that outsize gzip and bzip2's upper limits. Ppmz uses [Prediction by Partial Match](#) for compression, and has no upper limit on effective file size. PPM is a statistical model that uses [arithmetic coding](#). This gets us to things I don't really understand, and certainly can't explain here. Suffice to say that the authors found using ppmz that $C(x|x)$ always returned an NCD value between 0 and 0.1043. I looked around for quite a while and couldn't find a python implementation of ppmz, but I did find another [method ported to python](#) with [lzma](#), the compressor behind 7zip. Lzma uses a different implementation of Lempel-Ziv, utilizing a dictionary instead of a sliding window to track redundancies. What is more, the compression-dictionary can be as large as 4GB. You'd need a really, really large document to brush up against that. Though Cebrián, et al. didn't test lzma, my experiments show the NCD of $C(x|x)$ to be between 0.002 and 0.02! That's awfully close to 0, and the smallest return actually came from the longest document → 1781-1790.

THE CODE

In a way, that previous section is getting ahead of myself. I started with just zlib, and then added bzip2 and gzip, and eventually lzma for comparison sake. Let me clarify that just a bit. In python, there are two modules that use the gzip compressor:

1. gzip, which is for file compression/decompression; and
2. zlib, which is for compressing/decompressing strings or objects.

I was unsettled by my early zlib returns, and tried using gzip and file I/O, but got the same returns. Initially I was interested in speed, but reading Cebrián, et al. changed my mind on that. Nonetheless, I did time the functions to see which was fastest.

I based the script on Bill Turkel's back from 2007. (Bill put all of the scripts from the days of [Digital History Hacks](#) on [Github](#). Thanks to him for doing that!)

So, for each compressor we need a function to perform the NCD algorithm on a pair of files:

```
# Function to calculate the NCD of two files using lzma
def ncd_lzma(filex, filey):
    xbytes = open(filex, 'r').read()
    ybytes = open(filey, 'r').read()
    xybytes = xbytes + ybytes
    cx = lzma.compress(xbytes)
    cy = lzma.compress(ybytes)
    cxy = lzma.compress(xybytes)
    if len(cy) > len(cx):
        n = (len(cxy) - len(cx)) / float(len(cy))
    else:
        n = (len(cxy) - len(cy)) / float(len(cx))
    return n
```

There are small changes depending on the API of the compressor module, but this pretty much sums it up.

We need to be able to list all the files in our target directory, but ignore any dot-files like .DS_Store that creep in on OS X or source control files if you're managing your docs with git or svn or something:

```
# list directory ignoring dot files
def mylistdir(directory):
    filelist = os.listdir(directory)
    return [x for x in filelist if not (x.startswith('.'))]
```

Just as an aside here, let me encourage you to put your files under source control, especially as you can accidentally damage them while developing your scripts.

We need a function to walk that list of files, and perform NCD on every possible pairing, the results of which are written to a file. For this function, we pass as arguments the file list, the results file, and the compressor function of choice:

```
def walkFileList(filelist, outfile, compType):
    for i in range(0, len(filelist)-1):
        print i
        for j in filelist:
            fx = pathstring+str(i)
            fy = pathstring+str(j)
            outx = str(filelist[i])
            outy = str(j)
            outfile.write(str(outx[:-4]+" "+outy[:-4]+" ")
                +str(compType(fx, fy))+"\n")
```

That's all you need. I mentioned also that I wanted to compare execution time for the different compressors. That's easy to do with a module from the python standard library called `profile`, which can return a bunch of information gathered from the execution of your script at runtime. To call a function with `profile` you simply pass the function to `profile.run` as a string. So, to perform NCD via lzma as described above, you just need something like this:

```
outfile = open('_lzma-ncd.txt', 'w') print "Starting  
lzma NCD." profile.run('walkFileList(filelist, outfile,  
ncd_lzma)') print 'lzma finished.' outfile.close()
```

I put the print statements in just for shits and giggles. Because we ran this through profile, after doing the NCD analysis and writing it to a file named `_lzma-ncd.txt`, python reports on the total number of function calls, the time per call, per function, and cumulative for the script. It's useful for identifying bottlenecks in your code if you get to the point of optimizing. At any rate, there is no question that lzma is much slower than the others, but if you have the cpu cycles available, it may be worth the rate from a quality of data perspective. Here's what profile tells us for the various methods:

- zlib: 7222 function calls in 16.564 CPU seconds (compressing string objects)
- gzip: 69460 function calls in 18.377 CPU seconds (compressing file objects)
- bzip: 7222 function calls in 21.129 CPU seconds
- lzma: 7222 function calls in 115.678 CPU seconds

If you expected zlib/gzip to be substantially faster than bzip, it was, until I set all of the algorithms to the highest available level of compression. I'm not sure that's necessary or not, but it does affect the results as well as time. Note too that the gzip file method requires many more function calls, but with relatively little performance penalty.

COMPARING RESULTS

The Series Guide

A little bit more about the documents I'm trying to cluster. Beginning around 2002, the Archivo Nacional del Ecuador began to produce pdfs of their ever-growing list of Series Finders guides. The Criminales Series Guide ([big pdf](#)) was a large endeavor. The staff went through every folder in every box in the series, reorganized them, and wrote descriptions for the Series Guide. Entries in the guide are divided by box and folder (caja/expediente). A typical folder description looks like this:

Expediente: 6
Lugar: Quito
Fecha: 30 de junio de 1636
No. de folios : 5

Contenido: Querrela criminal iniciada por doña Joana Requejo, mujer legítima del escribano mayor Andrés de Sevilla contra Pedro Serrano, por haber entrado a su casa y por las amenazas que profirió contra ella con el pretexto de que escondía a una persona que él buscaba.

We have the place (Quito), the date (06/30/1636), the number of pages (5), and a description. The simple description includes the name of the plaintiff, in this case Joana Requejo, and the defendant, Pedro Serrano, along with the central accusation— that Serrano had entered her house and threatened her under the pretext that she was hiding a person he was looking for. There is a wealth of information that can be extracted from that text. The Series Guides as a whole is big, constituting close to 875 pages of text and some 1.1M words. I currently have text files for the following Series Guides—> Criminales, Diezmos, Encomiendas, Esclavos, Estancos, Gobierno, Haciendas, Indígenas, Matrimoniales, Minas, Obrajes, and Oficios totaling 4.8M words. I'll do some

comparisons between the guides in the near future, and see if we can identify patterns across Series. For now, though, it's just the Criminales striking my fancy.

The Eighteenth Century

So, what does the script give us for the 18th century? Below are the NCD results for three different compressors comparing my decade of interest, 1781-1790, with the other decades of the 18th century:

zlib:

```
cr1781_1790 cr1701_1710 0.982798401771 cr1781_1790 cr1711_1720
0.987881971149 cr1781_1790 cr1721_1730 0.977414695455
cr1781_1790 cr1731_1740 0.97668311167 cr1781_1790 cr1741_1750
0.975895252209 cr1781_1790 cr1751_1760 0.975088634189
cr1781_1790 cr1761_1770 0.975632632389 cr1781_1790 cr1771_1780
0.973381605357 cr1781_1790 cr1781_1790 0.974582153107
cr1781_1790 cr1791_1800 0.972256091842 cr1781_1790
cr1801_1810 0.973325329682
```

bzip:

```
cr1781_1790 cr1701_1710 0.954733848029 cr1781_1790 cr1711_1720
0.96900988758 cr1781_1790 cr1721_1730 0.929649194095
cr1781_1790 cr1731_1740 0.923066504131 cr1781_1790 cr1741_1750
0.906271163484 cr1781_1790 cr1751_1760 0.903237166463
cr1781_1790 cr1761_1770 0.902912095354 cr1781_1790 cr1771_1780
0.849356630096 cr1781_1790 cr1781_1790 0.287823378031
cr1781_1790 cr1791_1800 0.850331843424 cr1781_1790
cr1801_1810 0.850358932683
```

lzma:

```
cr1781_1790 cr1701_1710 0.965529663402 cr1781_1790 cr1711_1720
0.976516942474 cr1781_1790 cr1721_1730 0.947607790161
```

```
cr1781_1790 cr1731_1740 0.94510863447 cr1781_1790 cr1741_1750
0.931757289204 cr1781_1790 cr1751_1760 0.931757289204
cr1781_1790 cr1761_1770 0.92759202972 cr1781_1790 cr1771_1780
0.885106382979 cr1781_1790 cr1781_1790 0.0021839468648
cr1781_1790 cr1791_1800 0.880670944501 cr1781_1790
cr1801_1810 0.887110210514
```

First off, even just eyeballing it, you can see that the results from bzip and lzma are more reliable and follow exactly the patterns discussed by Cebrián, et al. The bzip run provides a $C(x|x)$ of 0.288, which falls in the acceptable range. The lzma run returns a $C(x|x)$ NCD of 0.0022, not much more needed to say there. And, as I noted above, with zlib/gzip we get 0.9745. Further, by eyeballing the results on the good runs, two relative clusters appear in the decades surrounding 1781-1790. It appears that from 1771 to 1810 that we have more similarity than in the earlier decades of the century. This accords with my expectations based on other research, and in both cases the further back from 1781 that you go, the more different the decades are on a trendline.

If we change the comparison node to, say, 1741-1750 we get the following results:

bzip:

```
cr1741_1750 cr1701_1710 0.888048411498 cr1741_1750 cr1711_1720
0.919398218188 cr1741_1750 cr1721_1730 0.826189275508
cr1741_1750 cr1731_1740 0.80795091612 cr1741_1750 cr1741_1750
0.277693730039 cr1741_1750 cr1751_1760 0.785168132862
cr1741_1750 cr1761_1770 0.803655071796 cr1741_1750 cr1771_1780
0.879983993015 cr1741_1750 cr1781_1790 0.906271163484
cr1741_1750 cr1791_1800 0.883904391852 cr1741_1750
cr1801_1810 0.886378259718
```

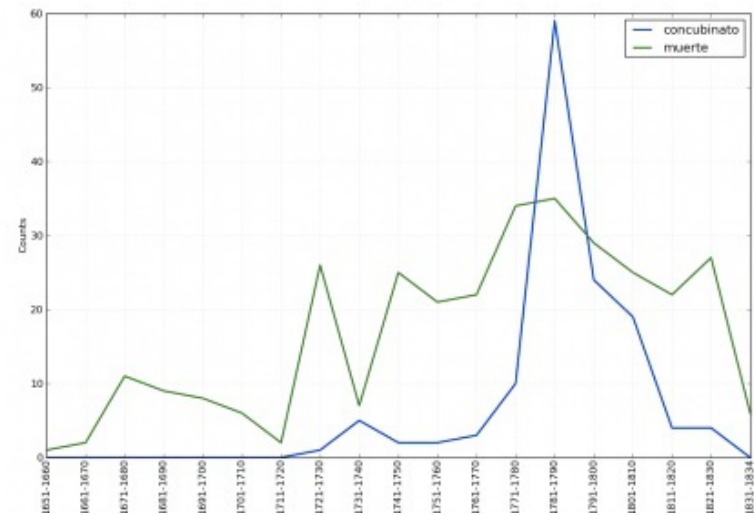
lzma:

cr1741_1750 cr1701_1710 0.905551014342 cr1741_1750 cr1711_1720
0.932600133759 cr1741_1750 cr1721_1730 0.862079215278
cr1741_1750 cr1731_1740 0.848926209408 cr1741_1750 cr1741_1750
0.00587055064279 **cr1741_1750 cr1751_1760 0.830746598014**
cr1741_1750 cr1761_1770 0.844162055066 cr1741_1750 cr1771_1780
0.90796460177 cr1741_1750 cr1781_1790 0.929573342339
cr1741_1750 cr1791_1800 0.908149721264 cr1741_1750
cr1801_1810 0.913968518045

Again, the C(x|x) show reliable data. But, this time bzip's similarities look a fair amount different than lzma when eyeballing it. I'm interested in the decade of the 1740s in part because I expect more similarity to the latter decades than for other decades in, really, either the 18th or the 17th century. I expect this for reasons that have to do with other types of hermeneutical screwing around, to use Stephen Ramsey's [excellent phrase](#) [PDF], that I've been doing with the records lately. Chief among those (and an argument for close as well as distant readings) is that I've been transcribing weekly jail censuses from the 1740s the past week and some patterns of familiarity have been jumping out at me. I have weekly jail counts from 1732 to 1791 inclusive, and a bunch others too. I've transcribed so many of these things that I have pattern expectations. And, the 1740s has jumped out at me for three reasons this week. The first is that in 1741, after a decade of rarely noting it, the notaries started to record the reason for ones detention. The second is that in 1742, and particularly under the aegis of one particular magistrate, more people started to get arrested than previous and subsequent decades. The third is that, like in the period between 1760 and 1790, those arrests were increasingly for moral offenses or for being picked up during nightly rounds of the city (the *ronda*). The differences are this—in the latter period women and men were arrested in almost equal numbers. There are almost no

women detainees in the 1740s. And, there doesn't seem to be an equal growth in both detentions and prosecutions in the 1740s. This makes the decade more like the 1760s than the 1780s. The results above bear that out to some extent, as distance measures show to be more like the 1760s than the 1780s.

I also had this suspicion because a few months ago I plotted occurrences of the terms *concubinatio* (illicit co-habitation) and *muerte* (used in murder descriptions) from the Guide:



Occurrences of the terms “concubinatio” and “muerte” from the Criminales Series Guide

You should see that right at the decade of the 1740s there is a discernible, if smaller, bump for *concubinatio*. I was reminded of this when transcribing the records.

CONCLUSION

OK, at this point, this post is probably long enough. What's missing above is obviously visualizations of the clusters. Those visualizations are pretty interesting. For now, though, let me conclude by saying that I am impressed initially to see the clusters that emerged from this simple, if profound, technique for clustering. Given that the distinctions I'm trying to pick up are slight, I'm worried a bit about the level of precision I can expect. But, I am convinced that it's worth sacrificing performance for either bzip or lzma implementations depending on the length of one's documents. Unless your files are longer than 900KB, it's probably worth just sticking with bzip.

Originally published by Chad Black on [October 9, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] Rudi Cilibrasi and Paul Vitányi, "Clustering by Compression," *IEEE Transactions on Information Theory* 51.4 (2005): 1523-45, [PDF](#).

[2] Ian H. Witten, Alistair Moffat, and Timothy C. Bell, *Managing Gigabytes: Compressing and Indexing Documents and Images*, 2nd Edition (San Diego, California: Academic Press, 1999), 21.

[3] Manuel Cebrián, Manuel Alfonseca, and Alfonso Ortega, "Common Pitfalls Using the Normalized Compression Distance: What to Watch Out for in A Compressor," *Communications of Information and Systems* 5, no. 4 (2005): 367-384.

[4] Cebrián, et al., "Common Pitfalls Using the Normalized Compression Distance," 372.

MARC DOWNIE AND PAUL KAISER

Spatializing Photographic Archives

The extensive and carefully illustrated White Paper for our NEH-sponsored "Spatializing Photographic Archives" project can be [downloaded as a large PDF](#) (26.5mb).

The White Paper describes the open-source software tool we've developed, and our reasons for wanting to forge a new approach to making digital tools for scholars. It also examines the implications of our approach for photography. After examining the history of landscape photography in the American West, we show how by stepping outside the photographic frame and unfreezing a photograph's frozen instant, we can reveal many hidden aspects of photography and create new kinds of works.

Our first case study investigates the Richard Misrach's canonical *Desert Cantos* series, which proved to be a difficult but exceptionally rewarding test case. In October 2009, we worked with Misrach at two of the original sites for the *Desert Cantos*.

At the first site, we reconstructed the ruins at the once-flooded edge of Bombay Beach on the Salton Sea in southern California, where there

remained enough landmarks for us to match our spatial reconstruction of the site to Misrach's original photos.



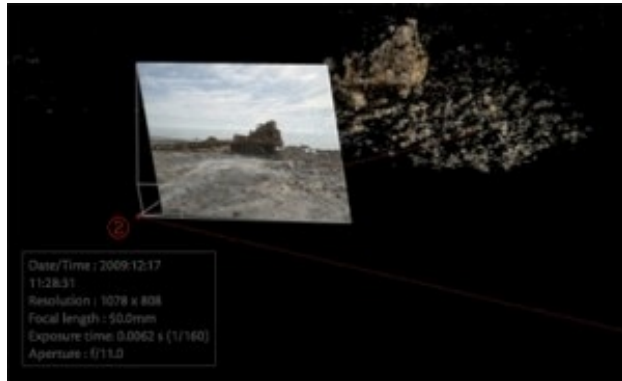
Reconstruction of the ruins at Bombay Beach

At the second site we spatialized a stand of palm trees that was the subject of several of his *Desert Fires* photographs.



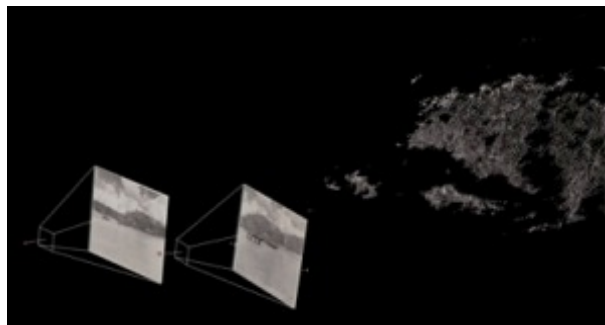
Spatializing palm trees from Richard Misrach's "Desert Cantos" series

We also reconstructed the process of one of Misrach's works in progress, spatializing his attempts to photograph a decrepit bulldozer at the edge of the Salton Sea. We track his path over the time of his shoot and his framing of the subject.



Spatializing Misrach's photographs of a bulldozer near Bombay Beach

The second case study examines battlefield photographs of Okinawa, 1945; the third prototypes a simple pipeline for scholars by which they make a 3D capture of an object using just the video capabilities of a smartphone and a laptop computer.



Spatializing of boats approaching the shoreline of Okinawa, Japan in 1945

Finally, the paper presents two hypothetical projects that our approach would underpin. These would create new kinds of interdisciplinary works that tie photo reconstruction to extensive data-mining, and would blur boundaries between the arts, humanities, and sciences.

Originally published by the OpenEndedGroup in [December 2011](#).

JEREMY BOGGS, ALISON BOOTH, DANIEL J. COHEN,
MITCHELL S. GREEN, ANN HOUSTON, AND STEPHEN
RAMSAY

“Humanities in a Digital Age” Symposium Podcasts

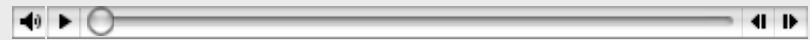
On November 11th, the University of Virginia's [Institute of the Humanities and Global Cultures](#) hosted a daylong symposium on “The Humanities in a Digital Age.” The symposium included two panels—one on Access & Ownership and the other on Research & Teaching—and two keynote talks.

The first keynote was given by [Stephen Ramsay](#), Associate Professor in the Department of English and Fellow in the Center for Digital Research in the Humanities at the University of Nebraska–Lincoln.

The second keynote was given by [Dan Cohen](#), Associate Professor in the Department of History and Director of the Roy Rosenzweig Center for History in New Media at George Mason University.

Originally published by the Scholars' Lab on [December 13, 2011](#). Keynote by Stephen Ramsay revised for the *Journal of Digital Humanities* March 2012 and available for download ([video](#), [PDF](#)).

JEREMY BOGGS AND ANNE HOUSTON



Panel 1: “Access and Ownership”

STEPHEN RAMSAY



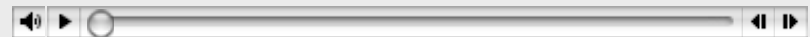
“Textual Behavior in the Human Male”

ALISON BOOTH AND MITCH GREEN



Panel 2: “Research and Teaching”

DAN COHEN



“Humanities Scholars and the Web: Past, Present, and Future,” with response by Jerome McGann

NIK HONEYSETT, AND MICHAEL EDSON

Philosophical Leadership Needed for the Future: Digital Humanities Scholars in Museums

Originally published as part of a [YouTube crowdsourced panel](#) for the Museum Computer Network Conference 2011 on the barriers to and benefits of implementing digital humanities methodologies in museums.

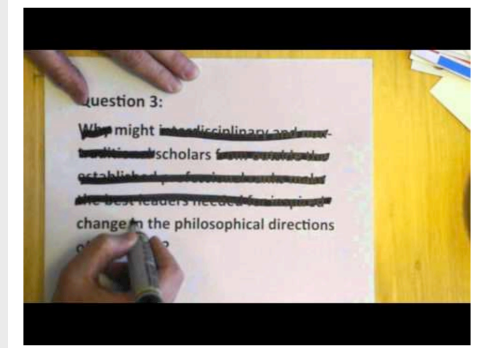
*Editor's Note: You must be connected to the internet to view videos

NIK HONEYSETT



Head of Administration for the J. Paul Getty Museum

MICHAEL EDSON



Director of Web and New Media Strategy for the
Smithsonian Institute

Critical Discourse in Digital Humanities

My interest in the role and nature of criticism in the digital humanities grows out of a question that Alan Liu recently asked: [Where is the cultural criticism in the digital humanities?](#) Although I'm not convinced that digital humanities needs its own brand of cultural criticism beyond what its constituents would normally do as humanists, the question resonated with me because it made me wonder (with only silence to follow): **where is the criticism in the digital humanities?**

Sadly, there really isn't any—a most unfortunate situation for both the innovative projects and people that constitute the digital humanities community. This essay explores the value of creating a critical discourse around scholarly work in the digital humanities. It's clustered around three main ideas:

1. Digital humanists have not created an effective critical discourse around their work.

I don't mean that we haven't gone far enough in publicly trashing each other's projects. Nor do I mean simply that there hasn't been enough peer review (though it's true; more on that later). Rather, I

mean that digital humanities criticism needs to go beyond typical peer review and inhabit a genre of its own—a critical discourse, a kind of scholarship in its own right—that will help shape best practices and make the value of our work clearer to those both inside and outside the digital humanities community.

2. We need more theoretical and practical rubrics for evaluating digital humanities work.

Although scholarly communication and peer review have been highly active topics in digital humanities circles for some time, these discussions have not yet produced truly useful evaluative criteria or discourse that appreciate the differences between digital and traditional work (not that there is a strict dichotomy). Everyone in the field knows that the most innovative digital humanities projects cannot be fully evaluated through the traditional, critical, and theoretical lenses of the humanities. But what lenses do we have? How do we know when to use them? How can we help others outside the field use them?

3. Digital humanities work requires a different kind of peer review to produce effective criticism.

The multifaceted nature of digital humanities work requires a different kind of critique than is typical in the humanities because it puts rather unique demands on both critics and criticism itself. Effective single-author reviews are almost an impossible expectation; useful critical discourse must be based on a new kind of collaborative and community-mediated peer review.

I: towards a critical discourse

We all know that disciplinary boundaries are notoriously difficult to define. Yet they remain recognizable beyond professional titles and departmental affiliations. This boundary problem also gives rise to the

question of whether there is any real difference between the humanities and the digital humanities—an interminable debate that need not detain us now. I don't believe that the digital humanities are fundamentally different from the traditional humanities in any larger epistemological sense, even if one takes the [hermeneutics of building](#) (often sloganized as "more hack, less yack") as a point of departure between them. It will suffice for present purposes to say that the digital humanities are—at least at the moment—*different enough* from the analog humanities.

Part of what defines a discipline is the rhetoric and aesthetics of its scholarly discourse. Philosophy texts sound different from history texts, which sound different from literary analysis. These differences become especially apparent during collaborative projects. As much as we champion cross-disciplinary work, there is an inherent unease to it, in no small part because it becomes more difficult to evaluate it. Given a particular piece of scholarship: How should one read it? Which criteria should be applied? Of course these lines in the sand are easily blurred and effectively dissolve if one looks too closely. But in the larger view, they exist and have consequences.

As disciplinary rhetoric and aesthetics help characterize and delineate different kinds of scholarly work, it's manifest to no small degree in the way work is evaluated by community consensus, convention, and ongoing efforts to codify practical and theoretical ideals. We have gotten very good at this in terms of traditional disciplinary work. One major way in which digital humanities is in fact separate from the humanities (again, at least for now) is that it requires new ways of evaluating very complex work in terms that are often unfamiliar to most humanists.

One of my favorite illustrations of the difficulty of evaluating digital work comes from William Thomas's article, "[Writing a Digital History](#)

[Journal Article from Scratch](#)." The article is from 2007, but describes events that seem almost ancient now, circa 2003. The article describes how analog historians critiqued history scholarship that did not look anything like the traditional journal article? Despite the project's many virtues, reviewers could only wonder what it did better than the standard practice, and whether "the rewards [of the website] were simply not commensurate with the effort and confusion involved." Well, it was a long time ago, you say. Agreed. But I'm confident that a similar exercise today would yield significantly similar results, certainly from a non-digital humanities audience, and—given the breadth of the digital humanities community—from plenty who identify as digital humanists themselves.

This is not to criticize the average humanist for not knowing the value of normalized data sets, relational databases, or valid XML. There is indeed more sensitivity to digital work, but the work itself has gotten considerably more complex as well. My point here is that those who *do* know their value haven't been particularly clear about why such technologies are useful in the context in which they're employed. What we might perceive as ignorance on the part of reviewers is at least in part because the rhetoric and aesthetics of digital humanities work is not particularly well established. In other words, the critical sphere has not yet materialized.

Why might this be? One reason for difficulty in fostering a critical discourse might center on the nature of the digital humanities community—a rallying point for many, if not most, self-proclaimed digital humanists. As a community, we've been encouraging and supportive, tending to include and welcome everyone with open arms. The "Big Tent" theme of the [2011 Digital Humanities Conference](#) suggests that it's ongoing. Such an approach has been essential and ultimately very successful in terms of broadening the

scope and influence of the field. This should, and hopefully will, continue.

However, such strong community solidarity and support may inadvertently curtail or discourage public criticism; the constantly expanding and amorphous boundaries of digital humanities itself further complicate evaluation. This is *not* to suggest that we should instead become exclusionary and inwardly hostile. But we can't be unhappy that tradition-bound humanists don't appreciate the value of our work when we haven't really outlined how it's different and how it should be appreciated. In other words, we haven't provided a public critical discourse that indicates to those without expertise what work is good and what is not—and thus serves as a compass for practitioners, critics, and outsiders alike.

Some post-talk tweeting prompts me to clarify two important points:

1. To argue for a critical discourse is not to suggest that digital humanities projects are inherently flawed as humanistic scholarship (a still far too common misunderstanding about the goals of digital humanities) and therefore must offer more tempered epistemological claims and greater transparency—and should be criticized when they don't. Good criticism will, of course, address these issues, but that's not really the point here. Criticism serves a much larger role beyond pointing out flaws (more on this later).
2. To suggest that there is an insufficient critical discourse surrounding digital humanities work does not inherently suggest that digital humanists have been generally uncritical in their work. We've all criticized projects, approaches, and results behind closed doors; we continually learn from and improve upon past work; we all think carefully about how to do our best work and sound

scholarship. But the most useful critical discourse is a public one. Exactly what constitutes the sound scholarship that we want to do (and actually do) is not nearly as apparent to others, especially those outside the digital humanities community, as it should be. It befalls the producers of that good scholarship to explain what is and what is not considered good, and why.

II. the value of digital humanities criticism

The long history of critical theory has well established the various functions of criticism, and need not be rehashed here. Digital humanities projects are not art, of course, and therefore may appear to have considerably less need for criticism, as opposed to simple peer review. I want to argue that a critical discourse of digital humanities work: (1) must be concerned with both interpretation and evaluation; (2) is central to establishing the importance of the kind of scholarly and even cultural work that it does. On the whole, a critical discourse will provide crucial services for an interested audience: establish utility and value, question blemishes and flaws, and identifies sources, commonalities, and missed opportunities. Criticism points out true innovation when it's perhaps not obvious that paint slopped onto a canvas is actually worth thinking about. It points out when success claims point to little more than—to adapt a phrase from Michael Joyce —technological frosting on a stale humanities cake.

Haven't we all seen intriguing, if not jaw-dropping, visualizations that made virtually no sense? Of course the real thrill of taking these in is to recognize the beauty that some obscene amount of data could be viewed in a small space, possibly interactively. Anyone who's even thought about creating visualizations from even well standardized data knows how difficult it is. The necessary technical triumphs notwithstanding, we need to discuss (for example) the value of being

able to automate the creation of such visuals apart from the communication that happens as a result of their design. We need to distinguish a methodological triumph from an interpretive one. Imagine how an explanation of the creation of such visuals could ease fears of black-box manipulation. This is just one instance where a critical discourse for digital humanities would be *far* more valuable than grant applications that sell potential work and post-facto white papers that champion whatever work got done. We need more than traditional journal articles that describe the so-called “real” humanities research that came out of digital projects.

Perhaps most importantly, as I’ve already suggested, criticism serves a crucial signaling function. [Matthew Arnold](#) in his *The Function of Criticism at the Present Time* defined criticism as “a disinterested endeavor to learn and propagate the best that is known as thought in the world.”^[1] One might easily and rightfully disagree with what critics like Arnold would have categorized as “the best,” but I think that his statement pretty well describes what we need to do. The staggering rate of digital humanities project abandonment has caused some alarm of late. One reason is that most academics aren’t that good at marketing beyond their disciplinary peers. Another reason is that it’s unclear what is worth emulating or learning from—especially for those new to the field. Criticism selects and propagates projects that deserve merit and to serve as models. To continue with the previous example, we need criticism that praises technological achievement of visualizing, while condemning poor design practices; we need criticism that lauds the interpretive potential while critiquing the potential transferability and reusability of the methodology.

Of course criticism has to be good and original, not dogmatic. [Irving Howe](#), the influential cultural critic from the mid-twentieth century, remarked that power of insight counts far more than allegiance to a

critical theory or position, as no method can give the critic what he needs most: knowledge, disinterestedness, love, insight, style.^[2] It’s not easy to have these! But criticism that reflects these talents performs extremely valuable scholarly work—work that goes far beyond the original project and makes it even more useful. Such criticism is especially good at establishing and debating terms of how to analyze a particular work. Discourse of critique is where new standards get hammered out. It’s the connective tissue of projects that pronouncements from on high simply cannot have.

One reason I’m evoking somewhat old-fashioned critics is because of the way in which evaluation and interpretation came together in their criticism. The way they hammered out a new kind of critique to judge, evaluate, and make sense of literature seem apropos to the new forms, structures, and processes digital humanities work, which is often fundamentally different than previous kinds of scholarship. For precisely that reason, it requires a different approach to evaluating and critiquing it. We need a critical discourse situated between contemporary critical theory that insists on interpretation and earlier schools of criticism in which evaluation played a much larger role. We need both.

The New Criticism of the 1940s and 50s largely relieved the critic from aesthetic debates about whether something should be judged good or bad. Critical emphasis refocused on interpretation of a self-contained object. This is not what we want to do with digital humanities criticism. Howe, a member of the “New York Intellectuals,” advocated connecting literary texts to their political and historical circumstances—an admonishment echoed by several pieces in the recent *Debates in the Digital Humanities*—though not aimed at particular projects—and such concerns should inform our public engagement with digital humanities work.^[3] The interpretive element remains important

because what the community can do with the results of a digital humanities project is, like art, often outside what a creator or project team might have envisioned for it—and this is where interpretation becomes important for multivalent digital humanities projects. What does it mean that a database has been structured in a certain way? What are the larger consequences for one design over another? How does a certain project push the boundaries of what we consider acceptable digital humanities work? How can new analytical processes or methodologies be applied in different contexts? These are subjective and interpretive questions that we must openly discuss.

As scholarship rather than art, the evaluative component that appealed to critics like Arnold must feature in any useful critical discourse. Contra Wimsatt and Beardsley, authorial intention must be considered, as formulated in Goethe's three questions for the critic: What was the aim? How well was it carried out? Was it worth it? Evaluative critiques are important because criteria for neither digital humanities work nor its evaluation have been well established. In broad terms, we might fruitfully follow literary critic Barbara Herrnstein Smith, who encourages us to evaluate according to integrity, coherence, boundaries, categories to which it belongs, and features, qualities, and properties that make it what it is.^[4] And we must do this not according to personal and subjective experience, but with respect to best practices as understood by the critic. A critical discussion of goals and outcomes will shape practices far more efficiently than decontextualized white papers or manifestos disconnected from implementations complete with their messy details and devil(s) lurking therein.

So what do we look for?

This last year in particular has seen much energetic rethinking of scholarly publishing and evaluative criteria for digital work. For

example, the MLA has outlined [types of digital work](#), as well as [guidelines for evaluation](#). To their credit, the MLA has been one of the most visible scholarly societies in starting and facilitating such discussions. However, here and elsewhere, the focus has remained on getting non-print work recognized and promoting the value of process over results.

These were important arguments to make (and to continue in some cases), but we must go beyond that now as well. Even if digital work is more acceptable, we haven't really created sufficient guidelines for evaluating digital work (broadly defined) on its own terms. Somewhat better in this regard are the [NINES guidelines for peer review](#), which call attention to usability and code. But overall these guidelines are at once too general to enable rigorous criticism, and too specific to NINES projects. In both cases, though, the suggested guidelines for evaluating digital work are not all that different from those for evaluating analog work. On one hand, that's exactly their point! On the other hand, it's perhaps a bit counter-productive because it doesn't sufficiently consider what's unique about digital work.

I'd like to outline a few very general criteria that might be broadly applicable to digital work, as disparate as it can be. I make no claims of completeness here.

Transparency

Can we really understand what's going on? If not, it's not good scholarship. "I used a certain proprietary tool to get this complicated visualization that took a gazillion hours to encode in my own personal schema—I won't bore you with the details—but here's what I learned from the diagram..." This cannot be considered good scholarship, no matter what the conclusions are. It's like not having footnotes. Even though we don't check footnotes, generally, we like to think that we

can. So it's natural to expect resistance when the footnote resembles a black box. Digital humanists have gained some traction in encouraging others to value process over product. Transparency helps us to evaluate whether a process is really innovative or helpful, or if it's just frosting.

Reusability

Can others export the methodologies, code, or data and apply them to existing or future projects? This embodies so much of what is central to the ethos of the community; we're always experimenting, always looking for better ways of doing things. The criterion of reusability creates an interesting gray area for generalized tools created by a project that aren't specific to that project, the humanities, or anything else—scripts that tidy up predictably messy data, for example. Are these to be shared and discussed? For now, absolutely! It's part of the effort, as Arnold suggested, to signal to our peers what proves most useful (hopefully, with justification). Obviously, not all project components or ideas are reusable. But discussion about what must be and what cannot be are important theoretical positions that will get worked out in a vibrant critical discourse about both concrete work and in abstract theoretical terms.

Data

Because most if not all digital humanities projects rely on data, it simply must be available—and not only for testing algorithms or verifying research results, but also for combining with other data sets and tools. Exactly how data should look is far from obvious. If nothing else, discussing a project's use of data will encourage conversations about interoperability, appropriate standards, ownership, copyright, citation, and so on. These issues are becoming more relevant than ever as we create new research corpora that bridge historically separate disciplines and archives. It's simply unacceptable to keep our data hidden—at least data that supports published work—with a wave of the

hand: “well, I cleaned this up and standardized it, and reformatted it ... but I'm going to keep this work invisible and hoard it.” It's like footnotes without page numbers.

Design

By "design," I really mean a project's underlying organizing principles. Our critiques must address why a particular design strategy is the best one or not, regarding both presentation and infrastructure. One challenge here is that academic convention dictates that—it least in terms of scholarly content—we privilege content over form. On one hand, good web design separates these; on the other hand, as McLuhan pointed out long ago: the medium is the message. Our attitudes about legitimacy and trust hinge on aesthetics in a relatively new way—almost a new kind of social contract between resource creators and users. Infrastructure is no less important, and perhaps less hidden than we suppose. Our critiques of design must also consider the decisions behind database design, encoding, markup, code, etc.

III: New kinds of peer review / criticism

My point here isn't just that digital humanities projects should embrace these values. Many already do. My point is that they need to get critiqued explicitly and publicly. But where and when do we do this? How does it fit with existing review and evaluative conventions?

As everyone knows, the nature of publishing has changed; we now do many digital projects that are never *really* done or officially published—at least not with an imprimatur of review and vetting. This means that the typical review process has been turned on its head. Getting a grant is too often an end in itself, taken to justify even the completed work. But this signing-off by the scholarly community happens *before* any work gets done. While traditional scholarship

(books and articles) is held accountable to its stated goals and methodologies (as far as the medium permits), digital projects have not had that accountability from the scholarly community. This is a grave disservice in two ways: projects learn less from each other, and projects remain isolated from relevant scholarly discourse.

It may sound as if I'm simply advocating for more peer review, as did a *Chronicle of Higher Education* article "[No Reviews of Digital Scholarship = No Respect](#)," which argued that scholarly societies and editors of traditional journals need to step up and encourage this work. Indeed, at least for now, peer review remains necessary for legitimization and certainly we should have more of it for digital humanities work. However, while change on the part of societies and journals would be nice, why should a few gatekeepers dictate the terms?

More importantly, I'm not convinced that getting a formal review and thus the imprimatur of serious scholarship is enough. We need a fundamentally *different kind* of peer review. Just as the nature of publishing is changing, the nature of peer review must evolve, especially for large, collaborative digital humanities projects, but even for small, individual ones as well. Digital humanities work requires a different kind of criticism than most academic criticism because of the very nature of the work. Digital humanities projects often serve much broader audiences and embody interdisciplinary practices in a way that eludes traditional models of critique.

I mentioned earlier the unease of situating interdisciplinary work in professional pigeonholes. As a way of fostering useful criticism, peer review needs be fundamentally collaborative in two ways:

- More people to review individual projects. How many people can really critique the various facets of a digital humanities project,

when they range from graphic design, interface design, code, encoding standards, etc. Even if one could, it's a herculean task not befitting the typical lone reviewer.

- Coordinated efforts to solicit and publish these critiques. Given the nature of complex publishing models of digital humanities projects, why not move away from editor-mediated peer review, which minimizes the public visibility?

So far I've discussed what happens after a digital humanities project is done. Ideally, projects could build into project timelines opportunities to solicit critiques, which hopefully will be feedback that avoids laudatory platitudes but rather shapes the project in productive, if challenging, ways. These might well be published as part of the project, which would help foster a vibrant critical discourse around the work being done. Good criticism will, of course, be applicable well beyond any particular project and constitutes scholarly work in itself.

Digital humanities work is often iterative in nature, and the review process needs to be as well. Just as digital humanities projects are inherently more public than the typical humanities project, everyone benefits when their critiques are more public. Project funders must prioritize and encourage public critiques as a way of establishing scholarly value and consider these critiques as part of funding decisions. A project without accountability, without connectedness, without critique, simply fills another plot in the digital humanities project graveyard.

One arena in which we might cultivate a vibrant and sustainable discourse is in the classroom. Our digital humanities courses need to explicitly teach critical methods for the unique issues in confronting digital humanities work. Both theory and practice is essential here. We must have more than gossipy complaints that don't

go beyond the classroom walls, or vapid reviews that fill the backs of most printed journals. Good criticism is very difficult. Students need practice pointing out what’s good and lacking in a project in a way that benefits both the project and the average humanist who needs to understand it. Such an environment makes it easy to get practice working in review teams that can leverage group expertise to address the variety of elements within a complex project.

IV: lastly



The criteria I mentioned earlier (transparency, reusability, data, design) operate in a larger theoretical context that we must consider as well. This might be profitably represented in an adaptation of a well-known diagram of criticism from M. H. Abrams, which shows four proximal spheres of criticism that might guide our approach. The formalist critique examines the form of the work itself, namely how well its structure, form, and design serve its purpose in the context of similar works. Didactic criticism focuses on the extent to which the work can reach, inform, and educate an audience. In the original

diagram, mimetic criticism would address the extent to which a work of art mirrors something larger about the world. For digital humanities criticism, it might evaluate how well digital humanities work accomplishes or facilitates humanistic inquiry. Lastly, the expressive critique discusses how well the work reflects the unique characteristics and style of the creator(s).

Of course these critical spheres are not entirely separate. In addressing each of them, for instance, we must remember that code and metadata, as well as data and whatever structures govern it, are not objective entities but are informed, attacked, and defended by ideology and theory. While these spheres of criticism might be applicable to humanities research generally, they are especially crucial for contemplating multifaceted digital work that is so often misunderstood. Furthermore, these different areas of critical focus provide opportunity for more critical theory in the digital humanities that grows out of its own work and also from further afield, drawing on critical methods from those working in new media and history of technology, as well as platform, hardware, and software studies.

Evaluative and interpretative public critique has much to offer the digital humanities. It's no panacea, of course. But it does seem like it could soothe the growing pains of a relatively new field or discipline or community, or whatever digital humanities should be called—especially since it depends on technologies and processes that change almost from day to day. But its problematic identity and transitory nature should encourage us to recognize the importance and power of a critical discourse—one that responds to particular projects—that will help explain, shape, and improve scholarship that demands new kinds of products and processes of engagement.

Originally published by Fred Gibbs on [November 4, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

This post is a moderately revised version of a presentation for [MITH's Digital Dialogues](#) series. This version has benefited from the thoughtful questions and comments that followed, as well as from the insightful critique of Natalia Cecire.

Notes:

[1] Matthew Arnold, *The Function of Criticism at the Present Time* (1864; New York: Macmillan, 1895), 81.

[2] Irving Howe, "The Conduct of Criticism: How We Do What We Do," *Salmagundi* 85/86 (1990): 55-70.

[3] Matthew K. Gold, ed. *Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2012).

[4] Barbara Herrnstein Smith, "Contingencies of Value," *Critical Inquiry* 10.1 (1983): 1-35; 18.

Conversations

Introduction: Theory and the Virtues of Digital Humanities

Natalia Cecire 44

When Digital Humanities Was in Vogue

Natalia Cecire 54

Theory First

Benjamin M. Schmidt 60

What We Think We Will Build and What We Build in Digital Humanities

William G. Thomas 63

Who You Calling Untheoretical?

Jean Bauer 68

Theory, Digital Humanities, and Noticing

Patrick Murray-John 75

Digital Humanities as Thunderdome

Elijah Meeks 79

Words and Code

Tom Scheinfeldt and Ryan Shaw 82

Please Write it Down: Design and Research in Digital Humanities

Trevor Owens 83

Building and Sharing (When You're Supposed to be Teaching)

Mark Sample 85

Marked Bodies, Transformative Scholarship, and the Question of Theory in Digital Humanities

Alexis Lothian 88

Where Are the Philosophers? Thoughts from THATCamp Pedagogy

Tim Sherratt 93

All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave

Moya Z. Bailey 102

Introduction: Theory and the Virtues of Digital Humanities

*I came to theory because I was hurting—
the pain within me was so intense I could not go on living.*

—bell hooks, “Theory as Liberatory Practice”[1]

The silicon chip is a surface for writing.

—Donna Haraway, “A Cyborg Manifesto”[2]

The debates around the role of “theory” in digital humanities are debates about the relationship between saying and doing. It therefore seems appropriately inappropriate to introduce a special section on digital humanities and theory with poetry, a kind of utterance in which language, it is still conceded, may *do* as well as *say*. Marianne Moore’s “In the Days of Prismatic Color” begins:

Not in the days of Adam and Eve but when Adam
was alone; when there was no smoke and color was
fine, not with the fineness of
early civilization art but by virtue
of its originality; with nothing to modify it but the
mist that went up, obliqueness was a varia-
tion of the perpendicular, plain to see and
to account for...[3]

The poem describes a prelapsarian world of unified meaning, in which “obliqueness was a varia-/tion of the perpendicular.” Once upon a time, the story goes, word and referent had a more than arbitrary relation, and the words “let there be light” could indeed call light into being. But this originary state of efficacious language met with a Fall, called “modernity.” In the beginning was the Word, but in the early modern period the Word devolved into mere “words, words, words.”[4]

With modernity, language’s relation to reality changed, and therefore so did the status of evidence. Gradually resemblance, discourse, and logical argumentation ceded epistemological authority to a factual register established through experimentation, witnessing, and testimony.[5] This was the seventeenth-century turn to “the experimental life,” as the historians of science Steven Shapin and Simon Schaffer termed it in their 1985 study *Leviathan and the Air-Pump*. Knowledge, once established through discursive proof, became a matter of the physical. We seem still to be in this modern moment. In the historical contest between the epistemologies exemplified by Thomas Hobbes’s *Leviathan* and Robert Boyle’s experimental air-pump, it seems clear that Boyle and the air-pump “won,” for a facticity resting on phenomena witnessed and recorded now sets the public standard for what counts as knowledge.[6]

But of course, it is Hobbes whose work we humanists are more likely to study and teach. The humanities’ perpetually defensive position vis-à-vis what Yeats rather contemptuously called “the noisy set/ Of bankers, schoolmasters, and clergymen” rests on this story of the modern Fall, for there is an archaic logic of resemblance that remains powerful and persuasive in humanistic inquiry, arguably underwriting its special ability to illuminate aesthetic questions. Indeed, if the humanities has ever seemed to triumph epistemologically, it was perhaps in the heady moment of “high theory” in the American university, when we learned

“how to do things with words,” and stories of man’s first disobedience seemed to have been—ever so briefly—undone.[7] Not for nothing were the 1970s and 80s the age of the argument by allegory and by pun.[8] But “It is no longer that,” as Moore puts it, and indeed, for many digital humanists this conceptual sundering of saying and doing—“hack” and “yack”—is not a thing to be mourned, but rather a *felix culpa*.

I am proposing, then, that the question of theory is a question about the place of digital humanities in a set of disciplines that have continually wrestled with the status of the word in the production of knowledge.[9] The essays included in this special section are therefore embedded in a complex set of institutional histories that bear on these questions of epistemological authority: the rise of the American university system, the relative cultural authority of the humanities and sciences within that system, the history of humanities computing, the “information age” and the close yet complex relation between knowledge and capital that characterizes it, the relatively sudden institutionalization of academic digital humanities, and its concomitant popularization—what Bethany Nowvskie has (with caveats) termed the “eternal September of the digital humanities.”[10] These histories are complicated and—of course—political. Gestures that consolidate professional legitimacy also name those actors who are and are not to be regarded as legitimate, with consequences that propagate unevenly across race, class, gender, sexuality, disability, and institutional status.

Epistemologies of doing

What, then, are the options for a postlapsarian humanities? One is to make the case publicly that “we have never been modern,” insisting on the mutual constitution of saying and doing.[11] This has been a powerful refrain in literary and cultural theory, from Wittgensteinian

language-games to Butlerian parody. By and large, however, digital humanities has taken another tack. Digital humanities does not so much contest the modern division between saying and doing as attempt to dilate the critical power of doing.[12] In its strongest version, digital humanities insists on an embodied, experiential, extradiscursive epistemology, what Jean Bauer, in her contribution to this special section, succinctly glosses as the assertion that “the database is the theory.” Historians of science often call such experiential knowledge “tacit knowledge,” but digital humanists generally call it “hacking.”[13]

The language of “hacking” pervades conversations around digital humanities, for instance, in Tad Suiter’s discussion of the term in his introduction to the crowdsourced volume *Hacking the Academy*. “Hackers,” Suiter writes, “are autodidacts. From the earliest hackers working at large research universities on the first networks to anyone who deserves the term today, **a hacker is a person who looks at systemic knowledge structures and learns about them from making or doing.**”[14] Stephen Ramsay—to some controversy—likewise promulgated an epistemology of doing in his remarks at the 2011 MLA panel on “The History and Future of Digital Humanities,” arguing that “if you aren’t building, you are not engaged in the ‘methodologization’ of the humanities, which, to me, is the hallmark of the discipline that was already decades old when I came to it.”[15] As Ramsay later elaborated, digital humanities is characterized by a “move from reading to making” that amounts to a fundamentally *nondiscursive* theoretical mode.[16]

For Ramsay, this epistemology of doing is necessarily a form of “tacit knowledge” that accounts for charges—like my own—that digital humanities is undertheorized. “At its most sneering,” he writes, “this is a charge of willful exogamy: we’re not quoting the usual people when

we speak. But there's frankly some truth to it." Ramsay goes on to quote Geoffrey Rockwell's argument that "[digital humanities] is undertheorized [in] the way any craft field that developed to share knowledge that can't be adequately captured in discourse is. It is undertheorized the way carpentry or computer science are." Happy fault—not only have *signs* lost the power to *do*, but *doing* has also lost its power to *signify*.

Epistemological claims are ethical claims

Suspending for a moment the question of whether this is necessarily the case—a question that Tom Scheinfeldt, Ryan Shaw, Trevor Owens, and Mark Sample take up in this volume—I wish to point out the ways in which these epistemological debates are implicitly ethical ones as well. We can already see the ethical dimensions of method in the rhetoric by which experimentalism came to be legitimated in the early modern period, as Shapin has detailed:

Experiments had really, and laboriously, to be done, not merely to be “thought.” [...] Rejecting traditional contempt for manual operations, the new gentleman-philosopher was not to think of himself as demeaned by mucking about with chemicals, furnaces, and pumps; rather, his willingness to make himself, as Boyle said, a mere “drudge” and “under-builder” in the search for god's truth in nature was a sign of his nobility and Christian piety. The rhetoric that presented new scientists like Boyle as craftsmenlike practical doers has been immensely effective....[17]

Boyle's experimentalism separated saying from doing, and made doing into a way to produce knowledge. In disarticulating saying from doing, the “experimental life” therefore reversed (but kept intact) the manual/mental hierarchy. This reversal was understood as an ethical good: the epistemology of doing was a repudiation of snobbery and an embrace of humility. Both rhetorics—of manual labor and of its ethical concomitants—are almost uncannily echoed in the disciplinary

discussions around digital humanities today. In addition to the language of “hacking,” terms abound that attribute to digital humanities a particular version of “doing” associated with manual labor: “hands-on,” “getting your hands dirty,” “dirt” (as in the Digital Research Tools wiki), “digging” (as in the Digging into Data Challenge), “mining,” and of course “building.”[18]

Perhaps the cleanest expression of the way that epistemological and ethical ideas travel in tandem is Tom Scheinfeldt's much-cited post on the “niceness” of digital humanities:

Digital humanities is nice because we're often more concerned with method than we are with theory. Why should a focus on method make us nice? Because methodological debates are often more easily resolved than theoretical ones. Critics approaching an issue with sharply opposed theories may argue endlessly over evidence and interpretation. Practitioners facing a methodological problem may likewise argue over which tool or method to use. Yet at some point in most methodological debates one of two things happens: either one method or another wins out empirically or the practical needs of our projects require us simply to pick one and move on.[19]

I am less interested in evaluating the claim than in bringing into relief the way that Scheinfeldt explicitly predicates a social relation—niceness—on the distinction between saying and doing (here rendered as theory and method) and, in particular, on the elevation of the latter over the former. Hacking is more than a method; it is an ethos.

Indeed, “niceness” is just one term in a whole set of ethical ground rules for digital humanities practices—what I have called, in my title, the “virtues” of digital humanities—which also include collaboration, humility, and openness.[20] Lisa Spiro has usefully codified some of these values, insisting simultaneously on their methodological and ethical valences, in her essay “‘This Is Why We Fight’”: openness (“on

several levels”), collaboration (“guided by a new *ethos*”), collegiality, diversity, and (harking back to Boyle) experimentation.[21]

Ethical claims are normative claims

It would be impossible, I think, to deny the salutary effects that this disciplinary epistemology-ethos has had on the wider profession; Korey Jackson seemed to speak for many when, shortly after the 2012 convention, he wrote that digital humanities was “how MLA found its heart.”[22] I am personally persuaded that digital humanists are almost universally committed to the ethical values that are emergent from the epistemology of doing—niceness, openness, and all the rest. And yet this ethos plays out in uneven ways, often with unintended consequences. Nowwiskie’s post on “eternal September” pointedly speaks to the ways in which compulsory niceness to noobs can lead to burn-out on the part of experienced digital humanists, and as Miriam Posner has more recently pointed out, “[s]ome people can easily afford to be nice; for others, the cost is higher.”[23] It is easier to be “nice” when one is not routinely met with casual racism, for example, and the costs of niceness—and of refusing to be nice—are distributed unevenly across race, gender, class, academic status and rank, and other social factors.

Who can afford to be a “hacker” or a “builder,” with the concomitant ethos of collaboration and niceness? In discussing Boyle’s self-presentation as an “under-builder” and “drudge,” Shapin and Schaffer observe that “it is absolutely crucial to remember who it was that was portraying himself as a mere ‘under-builder.’ Boyle was the son of the Earl of Cork, and everyone knew that very well. Thus, it was *plausible* that such modesty could have a noble aspect, and Boyle’s presentation of self as a moral model for experimental philosophers was powerful.”[24] It is not that Boyle was in any way disingenuous in

presenting himself as an “under-builder”—though many of his experiments were carried out entirely by the hands of servants in his employ—but that social factors positioned *his* “drudgery” as authorizing, whereas the literal drudgery of Boyle’s servants has meant their effacement from historical memory.[25] Certainly, no high school student today is taught “Boyle and his Assistants’ Law.”

The epistemology of doing, in a highly collaborative discipline often involving significant division of labor, means that, as labor is distributed across collaborators, so too is the attribution of knowledge. By this I do not mean “credit,” a much-discussed and serious question in its own right, so much as epistemological authority.[26] The manual/mental hierarchy, flipped in the valorization of “hack” over “yack,” too often returns in full right-side-up force just when it matters for *attributing knowledge to* the undergraduates hired to scan archival materials, say, or the workers in India who did the base TEI encoding. To espouse collaboration over authorship, one must have an authorial voice to cede; to be “nice,” one must be in a position in which “niceness” does not connote “servility.” Audre Lorde writes that “anger expressed and translated into action in the service of our vision and our future is a liberating and strengthening act of clarification.”[27] Does that “clarification”—a form of knowledge, to be sure—have a place in an epistemology of doing, with its ethos of niceness?[28]

Claims about doing are economic claims

In quite another register, the epistemology of doing has come to be framed in strangely specific terms, with social consequences for how it plays out in the wider discipline. “Hands-on,” “getting your hands dirty,” “digging,” “mining,” “building”—these terms offer quite a specific vision of what constitutes *doing*, conjuring up economic productivity (stimulus packages and infrastructure initiatives loom

into view) of a distinctly social, distinctly virtuous, distinctly white, male, blue-collar variety. The field might look very different if the dominant metaphors for “doing” digital humanities research included weaving, cooking, knitting, and raising or nurturing.[29] Indeed, we need not even look outside the academy for models for understanding the theoretical dimensions of praxis: *performance* and *activism* are forms of praxis that have been richly theorized in disciplines like performance studies, women’s studies, and ethnic studies. The (apparently accidental) choice of “building” for the dominant metaphor of digital *doing* was never an inevitability. In this postindustrial age, as Alexander Galloway has argued, “it is impossible to differentiate cleanly between nonproductive leisure activity existing within the sphere of play and productive activity existing within the sphere of the workplace,” and that feature of contemporary life appears to be one of the ones that the digital humanities, among humanistic subdisciplines, is uniquely equipped to handle.[30] Yet the distinctive methodologies of digital humanities are typically represented in comfortingly industrial terms.[31]

As Shapin and Schaffer write, “[s]olutions to the problem of knowledge are solutions to the problem of social order,” and indeed, to the still-beleaguered postlapsarian humanities at large, therein lies the great hope of the digital.[32] It is no secret that in the past few years many administrators have come to see in digital humanities a potential stimulus package for increasingly underfunded departments like English, history, comparative literature, classics, and so on. It is no wonder that alt-ac jobs, which require specialized skills and can be as difficult to attain as tenure-track jobs—or more—have come to be represented in the profession as shovel-ready projects just waiting to put our Ph.D.s back to work.[33] Digital humanities thus comes to be represented as a return to a (white, male) industrial order of union jobs and visible products, when in reality it is the subdiscipline of the

humanities most closely implicated in the postindustrial “feminization of labor,” with all that follows upon it: the rise of contingent and modular work, interstitiality, the hegemony of immaterial labor, the monetization of affect. Yet in its best version, digital humanities is also the subdiscipline best positioned to critique and effect change in that social form—not merely to replicate it.[34] In her essay “Theory as Liberatory Practice,” bell hooks recounts how she turned to theorizing as a way of “making sense out of what was happening.”[35] Surely such a making-sense is called for in this institutionalizing moment, and surely digital humanities itself is up to the challenge of doing it.

Claims make claims upon us

The question of “digital humanities and theory” ranges across historical, philosophical, institutional, and social registers, and each of the essays included in this special section attempts to address those registers in partial but interarticulated ways. The section begins with two posts that pose some initial questions, my own “[When Digital Humanities Was in Vogue](#)” and Ben Schmidt’s “[Theory First](#).” Contributions by [Will Thomas](#), [Jean Bauer](#), [Patrick Murray-John](#), and [Elijah Meeks](#) then consider the immanent knowledge in digital tools, whether as sources of theoretical richness or as undercurrents of unexamined assumptions. Brief comments by [Tom Scheinfeldt and Ryan Shaw](#) stake out strong positions in the question of whether digital humanities’s “tacit knowledge” demands to be rendered as discourse, whereupon [Trevor Owens](#) and [Mark Sample](#) expand on why and how digital humanists should aim to communicate their work to a wider public. The section ends with contributions by [Alexis Lothian](#), [Peter Bradley](#), [Tim Sherratt](#), and [Moya Z. Bailey](#). These pieces describe existing or imagined forms of digital “building” and “communicating” that benefit from explicit engagements with critical theory and its legacy.

One way of reading this special section might be as a soothing narrative in which the “provocation” of theory is raised, only to be shut down with the reassurance, in the end, that digital humanities is already “doing” theory, that no transformation is necessary, and that liberal “niceness” is already conducing to liberal equality. But I hope that readers of this special section will take it another way, as a serious questioning of the reluctance to “transform” despite our characteristic eagerness to “hack,” as a suggestion that we have only just begun to understand the ways in which “the database *is* the theory,” how we might “formulate a theory out of lived experience,”^[36] or the ways in which we might communicate “tacit knowledge” after all (say, to students who may not have had the luxury of developing their “tacit knowledge” by way of unlimited childhood access to a computer). Above all, I hope that the pieces we have included that suggest existing or imagined theoretical engagements for digital humanities will not be thought *sufficient*. The aim of this special section is not complacency but instigation.

As the quarterly journal stemming from the ongoing work of [Digital Humanities Now](#), the [Journal of Digital Humanities](#) selects online work in part on the basis of metrics that have shown that the work in question has *already* given rise to new thought and discussion within the field. This represents a response to recent calls for new, postpublication models of peer review. There are, of course, flaws in the system: for example, the group of scholars using the Twitter hashtag #transformDH, including Alexis Lothian, Amanda Phillips, Anne Cong-Huyen, Tanner Higgin, Micha Cardenas, Melanie Kohnen, and Anna Everett, have been central to the ongoing discussion of digital humanities and theory. Yet some of their most significant contributions have taken the form of face-to-face discussions, including at sessions at THATCamp SoCal and a roundtable at the 2011 American Studies Association conference. The online activity

generated by these formats is incommensurable with that generated by blog posts, and difficult to track. Similarly, because shorthand versions of who and what digital humanities is can self-reinforce in the social network, it is often difficult to catch work that expands our notions of the field’s boundaries. Moya Z. Bailey’s solicited contribution to this volume is meant to help counterbalance that centripetal tendency. Such examples show that as we work toward realizing a new model for peer review, structural gaps continually require our attention and correction.

As must by now be evident, I am not, for my own part, persuaded that the digital humanities’ epistemology of building is enough of a saving grace to render the hack/yack division a happy fault. My sympathies rest with bell hooks’s insistence that theory can solve problems that urgently need solving, that articulating in words how things work can liberate. I am troubled by the ease with which the epistemology of building occludes and even, through its metaphors, legitimizes digital humanities’ complicity with exploitative postindustrial labor practices, both within the academy and overseas, and I wish to see digital humanities dismantle as well as build things. And yet, as the contributions to this special section attest, the methods and metaphors of digital humanities are far from settled. What is needed is not self-flagellation (much less defensiveness) but attempts to develop the discipline within which we wish to work. This special section is offered to that end.

Acknowledgments

Many thanks to Joan Fragaszy Troyano, Dan Cohen, and the staff at PressForward for inviting me to work on this special section, and for their thoughtful collaboration throughout the editing process. Lauren Klein and Miriam Posner offered valuable feedback on this introduction as well. Some of these thoughts were developed in a

discussion with Stewart Varner’s graduate seminar “Topics and Tools in Digital Humanities” at Emory University; I thank Stewart and his students—Carla Almanza Galvez, Devin Brown, Anthony Cooke, Louis Fagnan, Scott Lisbon, Priyanka Sinha, and Tim Webber—for a lively and productive discussion. I am grateful to the Fox Center for Humanistic Inquiry at Emory University for research support.

Notes:

[1] bell hooks, “Theory as Liberatory Practice,” in *Teaching to Transgress: Education as the Practice of Freedom* (New York: Routledge, 1994), 59.

[2] Donna J. Haraway, “A Cyborg Manifesto,” in *Simians, Cyborgs, and Women: The Reinvention of Nature* (London: Routledge, 1991), 153.

[3] Marianne Moore, “In the Days of Prismatic Color,” in *Becoming Marianne Moore: The Early Poems, 1907-1924*, ed. Robin G. Schulze (Berkeley: University of California Press, 2002).

[4] Michel Foucault, *The Order of Things: An Archaeology of the Human Sciences* (New York: Vintage, 1994).

[5] Robert Boyle, Shapin and Schaffer argue, helped to bring about a radical epistemological shift in the early modern period. Whereas previously “knowledge” had been categorical, arising from logical demonstration, the English experimentalists established a form of knowledge that was “probabilistic and fallibilistic,” that is, contingent. Experimentalism’s physical “intervening” thus opened up a register of the factual that could compel assent partially and contingently, on the basis of observation, witnesses, and testimony, rather than on the basis of irrefutable logical or geometric demonstration. Demonstration

revealed knowledge “in itself”; experiment revealed knowledge by way of observation, testimony, and consensus. The turn to experiment was therefore also a turn away from the unity of signification that was demonstration’s premise—less a turn away from representation than the initiation of our postlapsarian sign/referent understanding thereof. See Steven Shapin and Simon Schaffer, *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (1985; Princeton: Princeton University Press, 2011), 23-4.

[6] It is this experimental legacy that has historically so deeply informed the digital humanities and, perhaps, stood in the way of its integration into the broader humanities. See Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Urbana: University of Illinois Press, 2011), 2.

[7] J.L. Austin famously describes a category of utterances that he names “performative,” before ultimately suggesting that all utterances are in some degree performative, and therefore that words do, indeed, *do things*. The moment of “high theory,” from Foucauldian discourse to deconstruction to Lacanian psychoanalysis, was deeply concerned to investigate the ways in which language could indeed *perform*. My comments above allude in particular to Jacques Derrida’s critique of the metaphysics of presence in *Of Grammatology*.

[8] In its ludic deformations of resemblance, Stephen Ramsay’s “algorithmic criticism” seems in sympathy with the deconstructive moment.

[9] Indeed, one of the great defenses of humanistic enterprises (if not what we now call “the humanities”), Shelley’s “Defence of Poetry,” is a response to Thomas Love Peacock’s satirical “Four Ages of Poetry” (1820), which argues in familiar terms that poetic discourse is irrelevant to an action-oriented modernity. In an important message to

the Humanist list in October 2010, Alan Liu argued that digital humanities was in a privileged position to undertake that defense because its aims and methods were congenial to a broad popular understanding of productivity. Liu's continued efforts in this vein have resulted in the work of 4Humanities, "a site created by the international community of digital humanities scholars and educators to assist in advocacy for the humanities," building on the digital humanities' "special potential and responsibility to assist humanities advocacy" ("Mission"). That digital humanities, with its thorny relationship to discursive methods, appears in this moment to be the best hope for "defending the humanities" perhaps brings into relief the deeper historico-philosophical issue at work. See Peacock, "Four Ages of Poetry" and Shelley "Defence of Poetry" in *Peacock's Four Ages of Poetry, Shelley's Defence of Poetry, Browning's Essay on Shelley*, ed. H. F. B. Brett-Smith (Oxford: Blackwell, 1953); "Mission," *4Humanities: Advocating for the Humanities*, accessed February 10, 2012, <http://humanistica.ualberta.ca/mission/>; Alan Liu to Humanist Discussion Group mailing list, October 25, 2010, no. 24.431, <http://lists.digitalhumanities.org/pipermail/humanist/2010-October/001653.html>.

[10] Bethany Nowviskie, "Eternal September of the Digital Humanities," *nowviskie.org*, accessed February 10 2012, <http://nowviskie.org/2010/eternal-september-of-the-digital-humanities/>. Nowviskie explains, "Eternal September' is a notion that comes from Usenet culture—the early peer-to-peer newsgroups and alt.* discussions that were, for many of us, an introduction to networked discourse and digital identity. Because Usenet activity centered around colleges and universities, a large influx of new students each September had disruptive effects on its established, internal standards of conduct, or netiquette. About thirty days in, newbies had either acclimatized to Usenet or they had dropped away, and the regular

roiling of September could be left behind for another eleven months. As the mid-1990s approached, Internet access became more common and less metered by the academic calendar. Once AOL began offering Usenet to its subscribers, September was eternal." Arguably, the recent institutionalization of digital humanities, with its concomitant proliferation of MLA and AHA panels, new digital humanities centers, NEH workshops, graduate certificates, and THATCamps, has produced a similar influx of "newbies" that, as Nowviskie observes, makes significant demands on the time, energy, and good will of experienced digital humanists.

[11] Latour frames modernity's great divide as a sundering of nature and the social, a version of the saying/doing divide. See Bruno Latour, *We Have Never Been Modern*, trans. Catherine Porter (Cambridge: Harvard University Press, 1993).

[12] A third view is endorsed by the *New York Times* columnist and erstwhile Miltonist Stanley Fish, who argues that we are indeed modern, and that the humanities are simply irrelevant (but, oddly, are nonetheless worth engaging in). See Fish, *Professional Correctness: Literary Studies and Political Change* (New York: Clarendon Press, 1995); Fish, "Will the Humanities Save Us?," *New York Times*, January 6, 2008, accessed February 15, 2012, <http://opinionator.blogs.nytimes.com/2008/01/06/will-the-humanities-save-us/>.

[13] As Lorraine Daston observes, there is no reason *in principle* that such knowledge should be "mute." See Daston, "On Scientific Observation," *Isis* 99.1 (2008), 101.

[14] Tad Suiter, "Why 'Hacking'?" in Dan Cohen and Tom Scheinfeldt, eds., *Hacking the Academy*, accessed February 12 2010, <http://>

www.digitalculture.org/hacking-the-academy/introductions/#introductions-suiter. Emphasis in the original.

[15] Stephen Ramsay, “Who’s In and Who’s Out,” accessed February 12, 2011, <http://lenz.unl.edu/papers/2011/01/08/whos-in-and-whos-out.html>. On the epistemology of building, see also Ramsay and Geoffrey Rockwell, “Developing Things: Notes toward an Epistemology of Building in the Digital Humanities,” in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 75-84.

[16] Stephen Ramsay, “On Building,” accessed February 12, 2011, <http://lenz.unl.edu/papers/2011/01/11/on-building.html>.

[17] Steven Shapin, “The Invisible Technician,” *American Scientist* 77 (November–December 1989), 556.

[18] On Digging Into Data, see *Digging Into Data Challenge*, accessed February 12, 2012, <http://www.diggingintodata.org/>. On DiRT, see Lisa Spiro, *DiRT Digital Research Tools Wiki*, accessed February 12, 2012, <https://digitalresearchtools.pbworks.com/w/page/17801672/FrontPage>.

[19] Tom Scheinfeldt, “Why Digital Humanities Is ‘Nice,’” *Found History*, May 26, 2010, accessed February 12, 2012, <http://www.foundhistory.org/2010/05/26/why-digital-humanities-is-%E2%80%9Cnice%E2%80%9D/>.

[20] We might understand collaboration as a methodological value rather than an ethical one, but in its usage it frequently has ethical overtones. Indeed, Roger Whitson argues that “collaboration is solidarity.” “Openness” is another ostensibly neutral term of art (as in “open access”) with an ethical dimension. See Whitson, “On DH, Solidarity, and Humility – or Why @miriamkp

Rules,” *rogerwhitson.net*, January 11, 2012, accessed February 12, 2012, <http://www.rogerwhitson.net/?p=1376>.

[21] Lisa Spiro, “‘This Is Why We Fight’: Defining the Values of the Digital Humanities,” in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: U of Minnesota P, 2012), 16-35.

[22] Korey Jackson, “‘Once More with Feeling’: How MLA Found its Heart,” *MPublishing Blog*, January 18, 2012, accessed February 12, 2012, <http://publishing.umich.edu/2012/01/16/mpub-mla/>.

[23] Miriam Posner, “Utopianism and Its Detractors,” *miriamposner.com*, January 11, 2012, accessed February 12, 2012, <http://miriamposner.com/blog/?p=1101>.

[24] Shapin and Schaffer, *Leviathan and the Air-Pump*, 66.

[25] On Boyle’s technicians and what they did, see Shapin, “The Invisible Technician.”

[26] The issue of credit is a separate and thorny one, on which see especially Bethany Nowviskie, “Where Credit Is Due,” *nowviskie.org*, May 31, 2011, accessed February 12, 2011, <http://nowviskie.org/2011/where-credit-is-due/>.

[27] Audre Lorde, “The Uses of Anger: Women Responding to Racism,” in *Sister Outsider: Essays and Speeches by Audre Lorde* (Berkeley: Crossing Press, 2007), 127.

[28] On refusals to be “nice,” see also Sara Ahmed on the “feminist killjoy,” in *The Promise of Happiness* (Durham: Duke University Press, 2010), 50-87.

[29] On “digging” in particular, see especially Bethany Nowviskie, “What Do Girls Dig?” *nowviskie.org*, April 7, 2011, accessed February 12, 2012, <http://nowviskie.org/2011/what-do-girls-dig/>.

[30] The emphasis on collaboration, collegiality, and the like in digital humanities bears striking resemblance to the values of corporate culture (e.g. “teamwork”) that Alan Liu describes in *The Laws of Cool: Knowledge Work and the Culture of Information* (Chicago: University of Chicago Press, 2004).

[31] Alexander R. Galloway, “Can the Whatever Speak?” in *Race After the Internet*, ed. Lisa Nakamura and Peter A. Chow-White (New York: Routledge, 2012), 120.

[32] Shapin and Schaffer, *Leviathan and the Air-Pump*, 332.

[33] See, for example, Leonard Cassuto’s recent call for Ph.D. programs to train graduate students for non-academic jobs. Cassuto quotes the public historian Richard Rabinowitz, who tellingly advises graduate students to “[g]et out of the library’ and ‘spend a year or two doing the work.” Cassuto, “Making a Public Ph.D.,” *Chronicle of Higher Education* (February 12, 2012), accessed February 17, 2012, <http://chronicle.com/article/Making-a-Public-PhD/130716/>.

[34] On the feminization of labor, see e.g. Antonella Corsani, “Beyond the Myth of Woman: The Becoming-Transfeminist of (Post-)Marxism,” trans. Timothy S. Murphy, *SubStance* 36.1 (2007), 107-138. See also “Devenir femme du travail et de la politique,” special section, *Multitudes* 12 (Spring 2003). On immaterial labor, see Michael Hardt and Antonio Negri, *Empire* (Cambridge: Harvard University Press, 2000) and Hardt and Negri, *Multitude: War and Democracy in the Age of Empire* (New York: Penguin, 2004).

[35] hooks, “Theory as Liberatory Practice,” 61. It is important to observe that hooks is deeply critical of theoretical work that, through elitist vocabulary and deference to the writings of white men, reproduce the structures of domination that they are meant to undo. Yet, as hooks writes, “[w]ithin feminist circles, many women have responded to hegemonic feminist theory that does not speak clearly to us by trashing theory, and, as a consequence, further promoting the false dichotomy between theory and practice. Hence, they collude with those whom they would oppose. By internalizing the false assumption that theory is not a social practice, they promote the formation within feminist circles of a potentially oppressive hierarchy where all concrete action is viewed as more important than any theory written or spoken.” See hooks, “Theory as Liberatory Practice,” 65-6.

[36] hooks, “Theory as Liberatory Practice,” 75.

When Digital Humanities Was in Vogue

“More hack, less yack,” they say. I understand the impulse, and to some degree admire the rough-and-tumble attitude of those in digital humanities whose first priority is getting things done. Hell, I like getting things done. But I cannot agree with the distinction between theory (little-t) and practice that this sets up, nor the zero-sum logic that it implies—i.e., that in order to do more we must speak less. “More hack, less yack” is, of course, just a slogan, a “spontaneous philosophy,” a stopgap. But stopgaps won’t do now that digital humanities is in vogue.

I mean for the title of this essay to refer to a line that Langston Hughes used to title a chapter of his memoir *The Big Sea* (1940), “When the Negro Was In Vogue.”^[1] Hughes’s ironic title frames an enduring and persistent philosophical and social question—race—as a matter of fads and fashions, “vogue.” In this, Hughes critiques the unintended consequences of the efforts of “race leaders” like W.E.B. DuBois and Jessie Fauset. Despite the real merit of black artists working in the period, Hughes suggests, the Harlem “vogue” seemed to get them into the spotlight on the wrong terms, laying the ground for a deeply problematic reception by the mainstream, whether through constant

comparisons to a bourgeois “white” artistic idiom or in a celebratory but ultimately dehumanizing primitivism.

So when I say that digital humanities is “in vogue,” I am talking about a new institutional prominence (i.e. in the last few years) that is only partially under the control of practicing digital humanists.^[2] This is what Bethany Nowviskie has guardedly described as the “eternal September” of the digital humanities:^[3] a new critical mass of digital work represented at major conferences like the Modern Language Association and the American Historical Association; new recognition of the need for standards for evaluating digital work for tenure and promotion; new digital humanities centers cropping up like mushrooms, with concomitant digital humanities cluster hires; the words “and digital humanities” suddenly ubiquitously tacked onto job ads; new grant opportunities; a proliferation of THATCamps. Consequent upon all of these are new burdens on the experienced digital humanists who have built the field. And one of those burdens—or perhaps I should say, responsibilities—is theoretical.

Given that digital humanists are now tasked with initiating much broader numbers of colleagues and graduate students into the field, how is that field to be represented? And what are the limits of a slogan in that pedagogy? Often, the new digital humanist is imagined as a fully formed humanities scholar who must now add some technical skills; thus, THATCamp workshops are usually dominated by computer-science-based technical skills or tools. The working assumption underlying this pedagogy seems to be that the “humanities” part of digital humanities is stable and more or less squared away, while the technical skills are what one needs to gain.

And again I turn to Hughes for a metaphor for teasing out the implications of that move. In *The Big Sea*, Hughes retrospectively satirizes those at the center of the Harlem Renaissance who “thought

the race problem had at last been solved through Art plus Gladys Bentley.” In the same way, now that digital humanities is in vogue, there is an overwhelming temptation to believe that the academia problem has at last been solved through the New Criticism plus code.[4] It’s the “plus” that makes Hughes’s comment so devastating: he puts his finger on a merely paratactic, additive concatenation that is the impoverished version of what can and should be a much more paradigmatic change. In other words, it should not be possible to have the “plus” without the two terms—“digital” and “humanities”—themselves changing.

Pedagogy is a place where we often oversimplify for the sake of clarity, even against our firmly held beliefs; this isn’t selling out—it’s good teaching. So perhaps it is natural that in the moment of the mainstreaming of digital humanities, much discussion of digital humanities remains characterized by that paratactic “plus.” The pedagogical emphasis on quick entry into the field—and the incredible success with which THATCamps, the Digital Humanities Summer Institute, and other initiatives have brought huge numbers of humanities scholars meaningfully into the orbit of digital humanities—is admirable. It also, however, comes with some costs. Theoretical keywords start to slide around in woefully unrigorous ways—words like “archive,” “labor,” “biopower,” “narrative,” “author.” You show up at a THATCamp and suddenly it seems that people are talking about separating form and content as if it were not only possible but unproblematic. The whole notion of “best practices,” pervasive in tech and industry, lives uneasily with theoretical critique. In taking up digital tools, it sometimes seems, we are asked to lay down our theoretical tools: more hack, less yack.

To be clear, I do not mean to caricature, much less insult, digital scholarship as it is currently practiced. The best digital humanities

work is already implicitly or explicitly theoretical, and in any case, there are times when you have to let a concept remain a black box if you are to do anything with it. Matthew Kirschenbaum has made the case for black boxes by proposing that “digital humanities” be understood as a “tactical” term, “to insist on the reality of circumstances in which it is unabashedly deployed to get things done—‘things’ that might include getting a faculty line or funding a staff position, establishing a curriculum, revamping a lab, or launching a center.”[5] Yet I want to press a bit on when those black boxes warrant opening, for the appeal to “get[ting] things done” (as opposed to theorizing) again reinscribes the particular black box I am here attempting to open. Is it not precisely in those moments of institutional incarnation that theory matters the most? As Kirschenbaum also reflects, “[o]nce a center is named, names are hard to change—who wants to have to redo the letterhead and the stenciling on the wall?” Strategic pedagogical oversimplifications take on new meaning in this institutionalizing moment, because they ramify, propagating and codifying themselves in new institutional structures.

Eternal September means that the theoretical commitments of digital humanities are more consequential than ever. And what are those commitments? “More hack, less yack” functions as a pedagogical shorthand because it really does capture something about the epistemological and ethical underpinnings of digital humanities. So what is that “something”? Or, to put it more bluntly, can such a representation ever be other than anti-intellectual?

To my mind, the best articulations of a digital humanities epistemology that rises above the shorthand have been offered by Stephen Ramsay, Geoffrey Rockwell, and Tom Scheinfeldt.[6] They have proposed that digital humanities is defined by immanent, nondiscursive modes of knowledge, which should be valued precisely in their nondiscursivity,

in ways analogous to performance or art practice or, in Rockwell's term, "craft disciplines."^[7] Haptic knowledge, intuition, know-how: these are real, if difficult and elusive, as anyone who has taught first-year composition (a form of so-called "yack") knows.^[8]

These critics have sought to elaborate the ways in which digital tools are theoretical tools. Rightly noting that writing is a practice that makes certain kinds of thinking possible, they propose an analogy with other constructive acts, notably the kinds of "building" characteristic of digital humanities research, which, they argue, demonstrates why digital building should be as recognizable as "scholarship" as writing is. This is persuasive, and fair enough as far as it goes. Yet Ramsay and Rockwell in particular go out of their way to defend against any contamination of the category of "building" (their translation of the informal "hack") by discourse.^[9] Ramsay, Rockwell, and Scheinfeldt give good accounts of why an epistemology of building need not be anti-intellectual—so long as that intellectualism is not overly given to discourse. And indeed, it entirely makes sense that these critics should attempt to isolate a form of knowledge that is not reducible to discourse, in order to investigate its status. But to then insist on its untranslatability seems to me to confuse the issue. Is it indeed necessary to strictly demarcate the construction of knowledge through writing (i.e. discursively) as different in kind from the construction of knowledge through (for instance) building a database? Especially if the latter's legitimacy as scholarship is being justified by the former's legitimacy as praxis?

These questions are prompted in part by a roundtable that took place at the 2011 meeting of the American Studies Association in Baltimore, which, to borrow Tara McPherson's pointed phrasing, asked why digital humanities are so white.^[10] I was particularly struck by part of the ASA roundtable description, which, without accusing anyone of

bad faith (and I agree; I don't think there is any), asks why the digital suddenly seems so congenial to the humanities just when ethnic studies departments and on-campus women's centers are getting axed (not to mention philosophy departments). The questions that the roundtable poses get at what we stand to lose when we fail to theorize practice, or when we insist on the tacitness of our theorizing:

In an era of widespread budget cuts at universities across the United States, scholars in the digital humanities are gaining recognition in the institution through significant grants, awards, new departments and cluster hires. At the same time, ethnic studies departments are losing ground, facing deep cuts and even disbandment. Though the apparent rise of one and retrenchment of the other may be the result of anti-affirmative action, post-racial, and neoliberal rhetoric of recent decades and not related to any effect of one field on the other, digital humanities discussions do often elide the difficult and complex work of talking about racial, gendered, and economic materialities, which are at the forefront of ethnic and gender studies. Suddenly, the (raceless, sexless, genderless) technological seems the only aspect of the humanities that has a viable future.^[11]

It is not so much that digital humanities is gaining at the expense of these programs (there's no direct correlation) as that something is making it easier to fund digital humanities just as it's getting harder to fund ethnic studies and queer studies. And so far, despite the best of intentions, digital humanities has not done a good job of theorizing either that disciplinary shift or its political implications. That's why I think we should probably get over that aversion to "yack." It doesn't have to replace "hack"; the two are not antithetical.

This brings me back to the Harlem Renaissance as a metaphor for digital humanities in the moment of institutional "vogue." Institutionalization seems to have prompted in the field the same sorts of identity crises that the Harlem Renaissance underwent. Despite

numerous essays on the subject, “What is digital humanities?” is the question we still constantly ask ourselves—not in the “I know it when I see it” way that we ask “what is modernism?,” but sincerely.[12] As Matthew Gold puts it, “[t]hese recent, definitional conversations bear the mark of a field in the midst of growing pains.”[13] Similar to the Harlem Renaissance, too, is the compulsive self-listing, self-mapping, self-visualizing, and general boosterism of (e.g.) totting up the number of digital humanities panels at this year's MSA, MLA, ASA, AHA, etc., comparing this year's number of digital humanities panels to last year's, comparing the MLA to the AHA, *und so weiter*. It reminds me of the lists of black writers in *The New Negro* and *The Crisis*—look how many we have! Have we not arrived?

And apart from Hughes and a few others, we see in the Harlem Renaissance a good deal of the target of Hughes's satire, Art plus Gladys Bentley—painfully derivative capital-A Art, glued to some of that Harlem vogue. In this volume, Fred Gibbs points out a good example of this in the phenomenon of “intriguing, if not jaw-dropping, visualizations that ma[k]e virtually no sense.”[14]

The comparison breaks down, of course. Digital humanities is not historically or substantively similar to the Harlem Renaissance, and in particular lacks the moral and political force of the Harlem Renaissance's sometimes misguided but deeply consequential efforts. But the way that the comparison breaks down is perhaps as important as the ways in which it holds. For one thing, it makes it all the more surprising when “the (raceless, sexless, genderless) technological” is rather unselfconsciously represented as somehow beleaguered in just the same way that women, the working class, and minorities have been. This is sometimes implicit in discussions of the difficulty of getting credit for digital work within humanities departments, which often bypass the ways in which it is significantly *easier* to get credit for

such work in mainstream culture (for instance, in the *New York Times*—I await breathless coverage of the latest in modernist studies) and, *a fortiori*, in the university as a whole, than it is to get credit for “traditional” humanities scholarship.

But the comparison occasionally even emerges explicitly, for example, in Mark Sample's borrowing of Milton J. Bennet's model of intercultural sensitivity as a metaphor for the stages toward acceptance of electronic literature, or in Nowviskie's suggestion that software development constitutes a “subaltern intellectual tradition.”[15] This is, I would argue, much too quick a shorthand for the real significance (ethical and otherwise) of the digital within humanities scholarship.

To note the internal tensions that the Harlem Renaissance and digital humanities share is to raise the question: why does digital humanities as a disciplinary formation—incongruously—seem to have so many ties in common with the Harlem Renaissance? What is the moral and political force of digital humanities—what are its cultural and institutional consequences? Are we content to suppose that it has no such force, or ought we not inquire?

Langston Hughes is right. Art plus Gladys Bentley is not going to get us where we're going, and the problem isn't Art, and it isn't Gladys Bentley—it's the plus.

Acknowledgements

This essay was originally published in a slightly different form as “When DH Was In Vogue; or, THATCamp Theory” on the blog *Works Cited* on October 19, 2011, and remains available at <http://nataliacecire.blogspot.com/2011/10/when-dh-was-in-vogue-or-thatcamp-theory.html>. Many thanks to Maria Cecire, Aaron Bady, and the staff at *JDH* for comments on earlier drafts of this essay. I am also

grateful to Ted Underwood and Sarah Melton for thoughtful comments on the blog post where this essay first appeared. I am indebted to the Fox Center for Humanistic Inquiry at Emory University for research support.

Originally published by Natalia Cecire on [October 11, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Notes:

- [1] David Levering Lewis adapted the phrase for his 1989 history of the Harlem Renaissance, *When Harlem Was In Vogue*. See Hughes, *The Big Sea*, introd. Arnold Rampersad (1940; New York: Hill and Wang, 1993), 223; Lewis, *When Harlem Was In Vogue* (New York: Penguin, 1989).
- [2] See Cecire, “It’s Not ‘the Job Market,’ It’s the Profession (And It’s Your Problem Too),” *Works Cited*, September 25, 2011, accessed February 29, 2012, <http://nataliacecire.blogspot.com/2011/09/its-not-job-market-its-profession-and.html>.
- [3] Bethany Nowviskie, “Eternal September of the Digital Humanities,” *nowviskie.org*, accessed February 10 2012, <http://nowviskie.org/2010/eternal-september-of-the-digital-humanities/>.
- [4] By “the academia problem,” I mean the long-term phenomenon that usually goes by the name “the crisis in higher education,” which has recently seemed to face accelerated rationalization, casualization, credential-inflation, and defunding. See “It’s Not ‘the Job Market,’ It’s the Profession (And It’s Your Problem Too).”
- [5] Matthew Kirschenbaum, “Digital Humanities As/Is a Tactical Term,” in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 415.
- [6] Stephen Ramsay, “Who’s In and Who’s Out,” accessed February 12, 2011, <http://lenz.unl.edu/papers/2011/01/08/whos-in-and-whos-out.html>; Ramsay, “On Building,” accessed February 12, 2011, <http://lenz.unl.edu/papers/2011/01/11/on-building.html>; Ramsay and Geoffrey Rockwell, “Developing Things: Notes toward an Epistemology of Building in the Digital Humanities,” in *Debates in the Digital Humanities*, 75-84; Tom Scheinfeldt, “Where’s the Beef? Does Digital Humanities Have to Answer Questions?,” in *Debates in the Digital Humanities*, 56-8.
- [7] Ramsay, “On Building.”
- [8] Lorraine Daston suggests another version of this kind of knowledge: “experience,” of the kind that allows a bird-watcher to identify a bird by its overall look and behavior at a glance. However, Daston argues, “[s]uch learned abilities are not in principle ‘tacit’; nor are they simply another expression of bodily skill, though they are also that.” See Daston, “On Scientific Observation,” *Isis* 99:1 (March 2008), 101.
- [9] Ramsay and Rockwell, “Developing Things,” 78, 81-3.
- [10] Tara McPherson, “Why are the Digital Humanities So White? or Thinking the Histories of Race and Computation,” in *Debates in the Digital Humanities*, 139-60.
- [11] Quoted in Alexis Lothian, “Conference Thoughts: Queer Studies and the Digital Humanities,” *Queer Geek Theory*, October 18, 2011, accessed March 15, 2012, <http://www.queergeektheory.org/2011/10/conference-thoughts-queer-studies-and-the-digital-humanities/>.

[12] Matthew Kirschenbaum's contribution to this genre remains invaluable, in my opinion. See Kirschenbaum, "What Is Digital Humanities and What's It Doing in English Departments?," in *Debates in the Digital Humanities*, 3-11.

[13] Matthew K. Gold, "The Digital Humanities Moment," in *Debates in the Digital Humanities*, x.

[14] Fred Gibbs, "[Critical Discourse in Digital Humanities](#)." *Journal of Digital Humanities* 1.1 Winter 2011.

[15] Mark Sample, "Electronic Literature Is a Foreign Land," *Sample Reality*, July 21, 2009, accessed March 4 2012, <http://www.samplereality.com/2009/07/21/electronic-literature-is-a-foreign-land/>. Bethany Nowviskie, comment on "Things We Share," *miriamposner.com*, March 5, 2012, accessed March 15, 2012, <http://miriamposner.com/blog/?p=1141&cpage=1#comment-31360>. The parallel here to what Timothy Yu has called "the ethnicization of the avant-garde" in the writing of Ron Silliman is striking. Yu documents how Silliman positioned Language writing as "marginalized" on same terms as women and racial, ethnic, and sexual minorities. Notably, as with the arguments made for digital humanities, this marginalization is understood as a matter of linguistic particularity. See Yu, *Race and the Avant-Garde: Experimental and Asian American Poetry since 1965* (Stanford: Stanford University Press, 2009), 38-72.

Theory First

It's easy to be reasonable about the relationship we'd like to see between digital humanities and "Theory." Each should inform the other. After all, humanists who put big-T Theory before any empirical data foolishly close their ears to the new evidence digital can create; digital humanists who ignore theory entirely jeopardize not only their careers but the soundness of their conclusions. To take two examples from the theory-friendly side of the spectrum in digital humanities; we should heed Natalia Cecire's [call](#) to treat digital humanities as important because it transforms humanistic practice; but we should also be mindful of Ted Underwood's concerns [that claims for the primacy of theory often amount to little more than a power play](#), serving to reify existing class distinctions inside the academy. In practice, this probably means digital humanists can keep calm and carry on, with greater tolerance for the occasional French name tossed into the discussion; meanwhile the theory-inclined should know they have a seat at the new table, though not necessarily at the head. Even more hack, better yack. What's not to love?

I've been flirting for a while with a much less reasonable point of view. It's based around two fairly tendentious convictions; both seem convincing enough to me that I want to try spelling them out.

1. Work in digital humanities should always begin with a grounding in a theory from humanistic traditions. If it doesn't, it will aimlessly reproduce a problematic social world.
2. The greatest hope for renewing our shared theoretical traditions in humanities research, and perhaps the only possible route, is to use massive stores of data digitally.

That is to say: theory and digital humanities aren't two separate enterprises that may be able to collaborate fruitfully. They are much closer to being one and the same thing. Digital humanities that doesn't put theory first ends up not really being humanities; social theory that doesn't engage with the explanatory power and communicative potential of vast digital data fails to take seriously its own conviction that deeper structures are readable in the historical record.

I've argued the second point elsewhere a bit, so let me focus on the first. (I should say that by theory, I mostly mean social or critical theory---those branches of philosophy that aim to change the world by understanding it. Just which one is not important here, though in practice, that is the only important thing.)

At their core, the digital humanities are the practice of using technology to create new objects for humanistic interrogation. (That's how I think of it, at least.) This has rightly led much of digital humanities' focus to lie in public humanities; there is enormous excitement about the potential of visualizations, exhibits, and tools to encourage non-humanists to think humanistically. ([I've talked about this before](#)).

But there is just as much reason to be excited about the prospects of creating new texts for humanists themselves to read. These are texts that bear little relation to the sort of books that we are used to reading. Visualizations, algorithmic rearrangements, and summary statistics aren't interpretations. They are texts in themselves. And they demand new sorts of mental gymnastics the same way that a newly discovered archive or poem does. The charts of the Stanford Literature Lab or the lists of Stephen Ramsay are creating new works that demand new kinds of readings; this development creates even more hope that digital humanities could transform the academic humanities at their core.

The trick is that we have to decide what new objects we want to read. Social networks, n-gram trajectories, interactive maps; objects that used to be prohibitively difficult to produce can now be assembled in an hour or a weekend. The technical chore of creating these new texts is neither as hard nor as important as figuring out what they should be. How do we decide what to make?

The answer, I am convinced, is that we should have prior beliefs about the ways the world is structured, and only ever use digital methods to try to create works which let us watch those structures in operation. The more scientifically minded might want to scream 'confirmation bias!' at this, but the wonderful thing about the humanities is that they have always allowed scholars to work from problem to evidence, not vice-versa. And while harnessing our work to theoretical agendas may dampen the ludic joy so easy to find in digital sandboxes, play alone can drift down dangerously well-worn paths.

The evidence and the tools at the disposal of digital humanists are not neutral. Research in the humanities has always been perilous, since our sources are so frequently shaped by those with power; digital proposes to do the same things to our tools. One of the things that I find the

most exciting about textual data is that for once we have a massive statistical store that wasn't collected by a state, with all the Foucauldian intimations contemporary historians are right to fret about. But without the agenda theory provides, we lose the distance from present power true criticism requires.

The unreconstructed texts of the past make us think in old ways. Archives, libraries, censuses, atlases: all of these force us to read juxtapositions far more aligned with historical ways of thinking than the reconfigurations possible with digital texts. Most historians, at least, are trained to think that this is fundamentally a good thing, because it gets us out of the cognitive ruts of the contemporary world. [The past is a foreign... something](#), and travel broadens the mind. I agree to a point that's good; nothing's more important for the historian than realizing that categories that are now sundered apart were once the same.

The promise and danger of the digital is that it lets us displace these texts, even though by only a hair's breadth, out of the systems of the past. Displacement is neutral in itself. Digital humanities would be a disaster if it simply rewrote our cultural heritage to fit neatly into the categories of the present instead of those of the past. That's why we need theory, which reconfigures the way we look at the world in terms of difficult to see structures that mask the truth: systems and lifeworld, doxa and habitus. There's a powerful significance there, and we need it.

The reason that digital humanities need to put theory first is not to pacify the powers-that-be, but to harness their own creativity towards productive ends. The solipsism of academia sometimes leads us to conflate power with tenure; but the real big game in the modern world does not wear tweed jackets. When humanists cite theory in protecting their turf, it is not just from luddism or self-regard; it is because they have a humane agenda, and fear that digital humanities do not. Some

of the great virtues of digital humanities---pragmatic usefulness, public outreach, borrowing from the sciences---only make it more suspect. Whatever the technical sophistication of digital humanities, it does not deserve to command those heights while its ends are impure.

Until then, skeptics are right to worry that all's not on the level. Something's fishy when a purportedly non-ideological movement shows up on the scene promising revolutionary change that looks suspiciously like the non-academic status quo. Why, exactly, should the 'next big thing' in the humanities come from the whitest, malest subfield this side of diplomatic history? Why does the *New York Times* cover the new field's projects so much more enthusiastically than it does traditional work? Why has digital humanities attracted more enthusiasm from state funders, across agencies and nation, than the humanities have seen since the Cold War ended? I often think: one of the things digital humanities is potentially very, very good at is naturalizing the world as it is. And our reflexive ways of thinking about the world are just what theory has always sought to get us away from; the nightmare from which it tries to jolt us awake.

Ted Underwood says that "[Theory](#)" is "[not a determinate object belonging to a particular team](#)." I'm not sure that's quite right. Theory belongs to all sorts of teams, but they all share something fundamental: they are the losers. The winners don't need new perspectives to shift their way of seeing from the world's; the losers do. What good the humanities have ever done largely lies in helping the losers along.

The digital humanities is perfectly poised at the moment to optimistically and beautifully affirm the world through all of history as it is now, full of progress and decentralized self-organizing networks and rational actors making free choices; or it might also try to take up what Adorno called the only responsible philosophy: to reveal the

cracks and fissures of the world in all its contradictions with otherworldly light. That's the demand placed on digital humanities by theory, and it must come first. [All else is mere technique](#).^[1]

Originally published by Benjamin Schmidt on [November 3, 2011](#).
Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] Theodor W. Adorno, *Minima Moralia: Reflections on a Damaged Life* (1951. London: Verso, 2005), 247.

WILLIAM G. THOMAS

What We Think We Will Build and What We Build in Digital Humanities

After five years "[workin' on the railroad](#)," I find myself confronting one of the central paradoxes of doing digital humanities--what Jerome McGann, one of the leading scholars of electronic texts, calls the problem of imagining what you don't know. In digital humanities, what we think we will build and what we build are often quite different, and unexpectedly so. It's this radical disjuncture that offers us both opportunities and challenges.

The [Railroads and the Making of Modern America](#) digital project, as it turned out, became my sub sub-library to borrow a phrase from Herman Melville and *Moby Dick*. And the work of the sub sub-librarian became one of classification and interconnection--it required getting out in the world too, talking with other collectors and librarians. In a way it required a different scholarly identity. Even so, as Melville warned, the archive "however authentic" offers only "a glancing bird's eye view of what has been promiscuously said, thought, fancied, and sung of Leviathan, by many nations and generations, including our own."[\[1\]](#)

When we produce a work of scholarship in whatever form, Jerome McGann reminds us that "to make anything is also to make a speculative foray into a concealed but wished for unknown." The work that we make, McGann tells us, "is not the achievement of one's desire: it is the shadow of that desire."[\[2\]](#)

I am particularly aware of McGann's disjunction right now, (and of Melville's caution), I suppose, because my project on Railroads and the Making of Modern America is at the end of five years. With the [Center for Digital Research in the Humanities](#), we have created a large digital archive, databases, visualization models, and some scholarly research publications. We have a cohort of graduate students in digital history trained and experienced. We have an audience of users.

But McGann's comment keeps raising its head. He tells us that that with which we conclude is only a shadow of the desired object. What we think we will build and what we build are not the same thing in digital humanities. We have only a "glancing bird's eye view."

This is as true of a book, a film, a painting, or a symphony as it is of a digital work. But right now, at this moment in the development of the digital medium, I think we can see how far we are from understanding the genre--of how far we are from being able to say send me "a prospectus" or its equivalent. The distance between our wish and our object is often so great because the forms and practices and procedures of creation in the digital medium remain profoundly unstable and speculative.

McGann's premise might be restated: if you have produced what you thought you would, perhaps you've not created anything really; if a digital project becomes what was specified it might not be a digital humanities work.

A series of questions have presented themselves, but what we really are asking in the broadest terms is how does scholarly practice change with digital humanities?

1. Is an archive an argument? And a related question, where is our scholarship?

Most projects in digital humanities begin as a digital archive, creating a collection of documents that are digitized. I want to encourage this—in the disciplines we need more attention to this work as scholarship. But digital scholars also seek to both assemble and analyze, both examine and interpret.

Five million books might be digitized, but the millions and millions of cubic feet of archival railroad records, well that was something else. What is a representative sample of railroad records?

We built a digital archive topically arranged for easy access and usability by the widest audience possible. Railroad texts were structurally so dissimilar that we confronted a major classification problem, one that we could not effectively address.

The architecture and encoding of a digital archive—what Johanna Drucker calls “creating the intellectual model” – must be undertaken speculatively.^[3] It must be adjusted, changed, explored. Interpretive archives cannot be built to spec.

Digital history has yet to fully confront the diversity of document types that we might wish to archive. We can build models from long runs of legal case files or printed texts or runaway slave newspaper advertisements, but when we turn to a domain such as railroads, or slavery, or genocide, or the family, the intellectual model behind an archive, so often expressed in encoded texts, becomes unwieldy. We

have *tended* to make archives of homogeneous document types, when the study of railroads, or slavery, or genocide demand much more capacious archives, with multiple, perhaps arbitrarily many, document types, as well as searchability across those types. In other words, we have tended to build archives that did not force us to confront these document-type problems, rather than the archives we truly need.

This challenge is our opportunity to reconsider the “digital archive” as intentional and interpretive—in our case to offer a new way to encounter the railroad. Rather than focus attention on the board room, or the directors, the archive can open up a diverse array of railroad users and interfaces. Its argument would be to expose the ways railroads were used and thought of. We want to create a new history of the railroad.

But as we create interpretive archives we need to be able to answer the question: Where is our scholarship? This is where we need allies—libraries in particular—as partners in modelling, preserving, and making available this scholarship.

The second question we face in digital humanities at this juncture is: How do we work differently?

2. How do we work in teams of scholars in the humanities?

Digital humanities projects are often characterized as collaborative. In many respects this is the most obvious change in scholarly practice—we work with librarians, programmers, and colleagues in other disciplines.

The opportunity here seems self-evident. But the model of historical and humanities scholarship has been sole-author, sole-researcher for a

long time, and for most universities the evaluation for hiring, promotion, and tenure proceeds to assess candidates on this basis.

In the Railroads project I wanted a team of graduate students to have the opportunity to gain experience in digital work, to advance their own scholarship, and where possible to participate in research publications. The challenge for digital humanities now is to make this work count where appropriate. We have begun keeping track of all research publications associated with the project—and we will be co-authoring new articles for the project with teams of researchers. In the early phase of digital humanities we built teams, and teams built projects. But now we are seeing teams contributing to publication streams.

The social structures for these contributions are not as yet settled. At the beginning of the project, I had only a vague idea how student colleagues would participate beyond building the digital project. Now, we are beginning to see projects build in publication objectives and contributions at the start.

A third question we face in digital humanities right now concerns the form of born digital scholarship.

3. What does scholarly argument look like in digital form?

My colleagues at the Center for Digital Research in the Humanities and my graduate students in the [Department of History](#) patiently bore with me on this one. From the first I hoped to experiment with a new form for our historical interpretative work, and this is what we began to call an “assemblage” or a “view.” The view is a framed set of materials on a given subject that integrates sets of evidence and data around a specific historiographical problem or question, without directly narrating the

subject. We wanted the views to inspire investigation and focus attention, to serve as interrelated starting points. We could have hundreds of views that build out of the collection.

The tools to assemble a view proved challenging to create—we were after all asking for an authoring tool for the digital medium. The rise of the blog in this same period reduced the incentive for experimentation with scholarly argument and hypertext.

The humble footnote is still the mark of scholarship and now we need to consider how we will migrate footnotes—the links and scholarly apparatus of a work—to digital form. This challenge and opportunity is surprising because the web is so good at linking. But we’ve not experimented as much as we could with discursive notes, linking, and narrative argument in digital form.

The changes in publication models should be an opportunity. We are on the cusp of a new genre of hybrid digital and print publishing. Books are and will be supported with digital sources and verifiable links to the elements that went into the study. Journals will move into the publication of born-digital work also, integrating print and digital formats.

In the humanities scholarly practice might shift toward a more fluid and open exchange of ideas and arguments characterized by a different sequence of activities:

- from openly available original research
- to pre-print presentation
- to peer review publication
- to a period of open verification

- to a period of adjustment and re-examination.

We know that opportunities and challenges here remain. We are in the early stages of this medium. We should look for ways to enchant readers, to hold attention, and to create long-form argument. Here we might be working against the medium (jumping through links) but the iPad and tablets appear to be opening up new opportunities for our scholarship.

Finally, we are in a transition phase. We call what we are doing “digital humanities” or “digital history” but really we are doing humanities in the digital age, we are doing history in the digital age. This work might be characterized increasingly by three qualities:

1. increasing the scale of research and data involved: 5 million books, 100,000 newspaper articles—this is the least important characteristic actually because it is limited to scholars, but the challenge will be not only to support this research with infrastructure but to come up with intellectual models for such large scale interpretation. Imagine how these “distant readings” fit in a U.S. history or literature survey.
2. addressing the global distribution of discourse and materials: sources all over the world need to be brought together and the challenge will be to create new linkages in the cultural records of the world, from Cairo to Seville to London to Chicago. Language differences, copyright, and sheer distance will need to be overcome.
3. using new models of production: we have students as colleagues and citizens as colleagues, and the challenge here will be to validate and credit their contributions, integrate their work, and do so in a way that enables further scholarship.

We are doing nothing less than redefining our practices and at the same time the relationship of our society to the past, our literature, history, and culture. Our digital age presents a different medium in which to convey multiple sources of information and to render interpretive arguments. It is instantiating different ways of knowing, different ways of seeing, reading, and learning. What we think we will build and what we build are not the same but we can and should celebrate and inquire into the difference.

Much has been made in our circles about Charles Joseph Minard’s [map of the Napoleonic March](#), but Minard drew his first such graphs for railroads in France and developed his technique in works combining traffic and distances. In 1845 he published what he called his first “figurative map.” Minard’s work, however, took more than 15 years to reach the sophistication we so admire. These 15 years witnessed the vast expansion of railroad culture in Europe and the U.S. Minard experimented with the forms for conveying multiple sources of information, but the disjunction between what he wished to build and what he built took time to resolve. We are, Robert Darton argues, perhaps in a similar position—15 years into what he calls the fourth great Information Age in human history. Like Minard, we are still learning how to adjust.

Acknowledgments

The above was originally given as a talk at the 6th Annual Nebraska Digital Workshop on October 14, 2011. I’m grateful to Kay Walter and Ken Price for the invitation to serve as a presenter at the workshop and to Susan Brown for participating on the panel, and to Kirsten Uszkalo, Jentery Sayers, and Colin Wilder for their participation in the workshop.

Originally published by William G. Thomas on [October 15, 2011](#).
Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] Herman Melville, *Moby Dick or The White Whale* (New York: Signet Classic, 1961), x.

[2] Jerome McGann, *Radiant Textuality: Literature After the World Wide Web* (New York: Palgrave MacMillan, 2001), 15.

[3] Johanna Drucker, *SpecLab: Digital Aesthetics and Projects in Speculative Computing* (Chicago: University of Chicago Press, 2009).

JEAN BAUER

Who You Calling Untheoretical?

[A note from the author: This blog post, as a piece of prose, is very much of the moment when it was written. Likewise its reception has been based on its tone as well as its content. So, rather than take this chance to revise the piece, I have decided to annotate the original text in the style of a documentary editor, although I have only annotated my own text, leaving the text of my commentators, Chris Forster and Jeremy Boggs (see below), alone. Aside from a few minor, silent corrections for editorial consistency, all new and supporting material can be found in the footnotes or set off by square brackets.]

[Providence, Rhode Island.

November 3, 2011 at 3:04pm]

I'm sorry. I need to vent. If you think you will be offended, continue at your own risk. You have been warned.[1]

Several weeks ago,[2] the whole "Digital Humanities Theory," or "Hack vs. Yack," debate sprang to life once more with a post by [Natalia Cecire](#). I have since read several other posts on this

issue, calling for more communication, [more give and take, more attention to political realities between Theory and Digital Humanities](#). [3]

However, I find many of the comments in these pieces insulting to those of us who work on digital humanities projects. I doubt this is intentional,[4] but I feel the need to defend the theoretical work already being done, while looking forward to incorporating even more ideas. Debate is good. In the academy, debate over terminology is inevitable yet often productive. So here is my rant:

I am sick and tired of people saying that my friends, my colleagues, and I do not understand or care about theory.[5]

Every digital humanities project I have ever worked on or heard about is steeped in theoretical implications AND THEIR CREATORS KNOW IT. And we know it whether we are classed as faculty or staff by our organizations. Libraries and other groups involved in digital humanities are full of people with advanced degrees in the humanities who aren't faculty, as well as plenty of people without those advanced degrees who know their theory anyway. Ever heard of [#alt-ac](#)? The hashtag is new; the concept is not.[6]

I have attended physical weeks of meetings to discuss terminology for everything from personal status (Do we label someone a "slave" or "an enslaved person?" If we have an occupations list should we include "wife," if so should we include "husband?" What about "homemaker?") to political structures (When do we call something an "empire?" Is "nation" an anachronism in this period?). I've seen presenters grilled on the way they display their index — and heard soul searching, intellectually rigorous justifications for chronological, thematic, alphabetic, or randomized results.[7]

Once I was presenting [The Early American Foreign Service Database](#) and got the question “So where is the theory in all of this?” Before I could answer with my standard, diplomatic but hopefully thought-provoking, response a longtime digital humanist[8] called out “The database is the theory! This is real theoretical work!” I could have hugged her.[9] When we create these systems we bring our theoretical understandings to bear on our digital projects including (but not limited to) decisions about: controlled vocabulary (or the lack thereof), search algorithms, interface design, color palettes, and data structure. Is every digital humanities project a perfect gem of theoretically rigorous investigation? Of course not. Is every monograph? Don’t make me laugh.

I have spent so much time explaining the theoretical decisions underlying [Project Quincy](#), that I wrote a program to allow database designers to generate color-coded, annotated, interactive database diagrams in the hopes that more Humanist Readable documentation would make all our lives easier. (The program is called [DAVILA](#).)

One of the most exciting things about digital humanities is the chance to create new kinds of texts and arguments from the human experience. Data structures, visualizations, search tools, display tools . . . you name it . . . are all a part of this exploratory/discovery process.

So it’s time for me to stop ranting and, in the best digital humanities tradition, DO SOMETHING.

If we as digital humanists are creating something new, then I believe our vocation includes teaching others how to read our work. If someone looks at The Early American Foreign Service Database and doesn’t see the theory behind it, maybe I need to redesign the site. Maybe those color-coded, annotated diagrams should be more

prominently displayed. Maybe I need a glossary for my controlled vocabulary. I wrote DAVILA, but the download only parses one kind of schema. Maybe I should write some more.

I’m going to stop talking (for now.) But, I’ll end with a tweet from [Matthew Kirschenbaum](#), a great practitioner and theorist of digital humanities: “More hack, more yack, and please, cut DH a little slack. We’re just folks doing our work.”

[Keep reading for the excellent comments on the original post.]

[The comment thread begins here.]

[Chris Forster](#) says: November 4, 2011 at 12:04 am

Hey Jean, as a (sort of) former colleague and current friend, thanks very much for this post. It has crystallized for me a key sense of where this tension between hacking and yacking is coming from.

I’d start by noting the slippery grammatical place of the word “theory” in this post. To people’s desire for “theory” you answer with comments about matters “theoretical”—“theoretical decisions,” “theoretical understandings,” or “theoretically rigorous investigation.” That is, I take you to you understand “theory” to mean something like “the (deeper) understanding of our material/objects of research” which necessarily undergirds our practice as scholars.

To many, however, “theory” means something related but more specific; it’s really just shorthand for a selection of French writers (Derrida!). Certain Germans can be “theory” in a pinch (Adorno, Kittler, Nietzsche, Hegel); even, once in a great while, an American will make the cut (Cavell, Butler)

I'm being funny (or trying); but I'm not really kidding. Indeed, when people wonder about the hack/yack ratio, I actually think they're asking a productive and important question about how current research relates to some of the best established traditions of humanistic research over the last three decades.

With that in mind, let me try two points:

The appeal to theory is often an appeal to connect digital humanities to a recognizable tradition in the humanities: at a number of [Scholars' Lab](#) talks my pet question, often carefully placed in the mouths of "my non-digital humanities colleagues in the English Department," would be "Yes, but how does this help me study the history of sexuality?" I would often be assured that, indeed, it could. But at that very moment I think I may have been not expressing myself clearly. The real question I was trying to ask was something like: can you talk a little about the repressive hypothesis and how your method would approach or complicate the account offered by Foucault in *The History of Sexuality*. You can substitute another proper name and another question, because really, beneath this, I was just asking to be talked to in a vocabulary I recognized. Part of "Theory's" importance, despite everyone's gripes about it (I've got your jouissance right here!), has to do with its ability to allow people in different fields to talk to one another. A Victorianist, a modernist, and a medievalist can all talk about gender performance. (Or, at least, that's the fantasy.)

This feels especially pressing now because certain (high profile) projects often seem to have a naive theoretical grounding: "naive empiricism" or "mere positivism" are the sorts of objections one hears, particularly around projects which are using digital methods to examine large amounts of data. These projects, and the statistical methods which many of them rely upon, engage the autoimmune reaction of many humanities scholars who have strong reactions to

anything smacking of empiricism. The reaction to Google nGrams I think captures this.

People smarter than me will point out that digital humanities "building" is itself a theoretical activity; that the theoretical roots of digital humanities start in Plato and pass through Heidegger and continue through folks thinking about textual materiality (e.g. Kittler, McGann, and Kirschenbaum). Yup. And I think this is the very terrain on which discussions between theory and digital humanities might begin or, more properly, continue.

Another response to what I'm saying is: "So what? Zizek/Derrida/Latour/Levinas is not important to my project. What is important is not 'Theory,' but theory—not genuflection before the idols of the past, but rigorous self-interrogation of our method." And that seems fair enough as far as it goes. But I worry it dismisses too quickly texts which have proved important to many folks calling themselves humanists.

There is, I think, a real debate about method, value, and purpose in the humanities which is expressed by tension over the hack/yack ratio. I would try to take the request for "more theory" in digital humanities not as an insult or an accusation, but as a serious invitation to a conversation. Just as "non-digital humanists" should resist the criticism that digital humanities is a just a funding-hungry, shiny-tool-obsessed attempt to reduce cultural study to word frequency histograms, "digital humanists" should likewise resist the sense that "theory" is just a code word for "the same old same old," coming to grind the gears of hackery to a navel-gazing, yack-yack-yackety stop.

Hope all's well up North Jean; we miss you on this side of the Mason-Dixon.

[Jean Bauer](#) says: November 4, 2011 at 11:49 am

Chris: Many thanks for the thoughtful response. This is precisely the conversation I wanted to have. As I hope I made clear before the rant began, I am all for debate. More ideas, more discussion, all to the good.

Let me start by saying that, as usual, I basically agree with you. Moreover, I really appreciate the clarification you provide. As you said, the “slipperiness” of language is (in part) at fault, but I am not the only one who is having this problem. In my humble opinion, people who are using the word “theory” to stand in for a specific set of writers and ideas, which have permeated different disciplines and sub-specialities in the humanities to different degrees, might want to keep that in mind as well. Digital humanities is a big tent and that means a lot of time spent defining our terms and possibly creating some new ones (which will in turn need to be defined;-)

What I was really ranting about, though, are some types of comments that seem to keep popping up in these posts which, regardless of initial intent, I find hard to read as anything other than insulting. I am referring to things like “using the Author field in [Drupal](#) uncritically” or discussions of theory as a “power grab” by tenured faculty against staff. Many of us think very critically about the tools we use — often choosing to make our own tools and schemas rather than work in systems we find to be theoretically and/or Theoretically insufficient. And you do not have to be faculty to have read Derrida. Some faculty might believe that, but they are wrong. And depending on your field of study, the faculty may not read Derrida.

I try not to be insulted. But I also believe that civil discourse occasionally requires someone standing up and saying “Hey! That was really insulting.” Then we can all sit down, unpack our terms, and go from there.

[Jeremy Boggs](#) says: November 5, 2011 at 11:28 am

Thanks, Jean for a great post. And thanks, Chris, for a terrific response to it. If you don’t mind, Chris, I’d like to explore some of my own feelings on all of this, since you touch on a few things that generally bother me about this whole thing.

To use your Foucault example, it seems perfectly reasonable to me for someone to build some kind of digital humanities project that has nothing to do with the repressive hypothesis, or how it might complicate Foucault’s account in *The History of Sexuality*. And that same person shouldn’t necessarily know how their work might complicate Foucault’s account. I would dismiss this because I don’t necessarily feel it’s my job to explain how the digital humanities project I’m building, or the methods and technologies I’m using, will help them do their work better. It’s my job to explain why I took the approaches I did, certainly, which is itself something lacking in digital humanities. But, I feel like it’s their job to critique my project as it is, and then to discuss how it might impact their work, if at all. I do however, feel it’s a great opportunity for you, or whoever, to talk about that in some form, or at least explore it further. If it doesn’t, then we all move on.

Another response to what I’m saying is: “So what? Zizek/Derrida/Latour/Levinas is not important to my project. What is important is not ‘Theory,’ but theory—not genuflection before the idols of the past, but rigorous self-interrogation of our method.” And that seems fair enough as far as it goes. But I worry it dismisses too quickly texts which have proved important to many folks calling themselves humanists.

I feel like this is my usual response to most of these points about theory, and I readily admit that I probably dismiss them too quickly.

But my response would be more like “How do you think Zizek/Derrida/Latour/Levinas would complicate my argument or project?” The fact that a digital humanities project doesn’t take into account a particular theory, or theory in general, is not at all a failing on the part of DH. This is like saying that because a scholar doesn’t take an approach I think is valuable, their work is no good. It doesn’t actually engage or critique the author’s own argument or methods. “The project doesn’t do what I want it to do, so it’s lacking in some way.” (I detest this kind of response to any scholarly work.) This is how I feel about most of the recent calls for more theory in digital humanities, and this just feels silly to me. I don’t really feel insulted by any of this; I feel unimpressed.

But, I don’t want to feel unimpressed at all. I love debate, and I love learning new things and exploring new approaches. I’m open to seeing more theory in digital humanities, but I’d like to see some folks actually do that, instead of talking about doing it, or criticizing digital humanities for not already doing it. They should just start doing it, at every THATCamp they attend, or on blogs, or wherever possible.

So am I missing something? Are people already doing this, and I’m just missing it? Is there something wrong with my reaction? Something I’m overlooking or ignoring?

[Chris Forster](#) says: November 7, 2011 at 9:30 pm

Jean: I think you’re right that “Theory” has “permeated different disciplines and sub-specialities in the humanities to different degrees,” and that this may be a source of unintentional confusion. As is often the case, part of the difficulty we (collectively; not you and I, of course!) have in communicating has nothing to do with the digital but everything to do with the humanities. The way I describe capital “T” Theory may be more peculiar to literary study, where courses which

begin with the nineteenth-century trio of Nietzsche, Freud, & Marx, and trace a path through structuralism, psychoanalysis, and figures like Derrida, Lacan, and Foucault, are in many universities (even for undergrads). These courses are often serve as a *lingua franca* (or, perhaps, merely a Frenchified argot) across fields within literary studies—a function they may not serve in other disciplines. Such courses, in fact, are often what folks in literature departments consider methodology.

There is a point to be made here (and others, like McGann, have made it) about a tension within literary studies, and the way that interpretation has come to dominate literary studies—as opposed to other modes of scholarship on literature, like textual criticism, scholarly editing, philology, etc. And so part of what my question about Foucault and *The History of Sexuality* is asking is, how does digital humanities change how I interpret texts; this is a different question, I think, than how do digital technologies change I understand the past (the historian’s question?). Because, to some extent, “interpreting texts” seems like a fundamental part (maybe the fundamental part) of being a literature scholar.

This perhaps lets me say something to Jeremy’s response, which I especially appreciated because I think it points to what I think is a genuine sort of miscommunication or misunderstanding. Jeremy writes, “I don’t necessarily feel its my job to explain how the digital humanities project I’m building, or the methods and technologies I’m using, will help them do their work better.” I think its relevant that who “they” are is not entirely clear from the context. Your point seems absolutely fair and, as you say, seems at least as true of research projects and agendas which aren’t under digital humanities’ (even very capacious) big tent.

But the scholar I imagine asking about *The History of Sexuality* is not asking for someone else to do his/her work. He or she is asking: what can we talk to each other about? They are asking even, "Are you talking to me?" (ideally not like [this](#)). Admittedly, in the culture of academia this question often has an edge and we may sound more like Travis Bickle than we should. I too find something frustrating about this sort of question when it degenerates (as it often does at, say, academic conferences) into: "yes, but why didn't you talk more about my topic?" This is, I think, a tolerable evil of trying to talk to one another. This perspective is, I'll readily admit, naive about the politics of how universities are organized. But in the conversation about "theory" and digital humanities, I do hear a genuine question not simply a power play.

Let me end by being specific and mentioning two instances of people doing theory-infused digital humanities or perhaps digital humanities-infused theory. I don't know that either would appreciate this designation and so I offer merely my perspective on my limited sense of these people's work: I've only seen Jo Guldi speak once (at, of course, the Scholars' Lab; hear it [here](#)); but what impressed me most was how seamlessly she embedded new digital methods in an existing critical discourse (by Jove, Foucault's in there!). Here "Theory" establishes something pretty basic—an existing scholarly discourse. Digital humanities projects are always humanities projects; but here Jo does a remarkable job of making that link clear.

The other is perhaps more apropos to our discussion; [this essay](#) by Johanna Drucker (recently discussed by some folks at the University of Virginia as part of the EELS group (electronically enabled literary studies)). It represents a critique of sorts of what Drucker claims are the danger of visualization in the humanities. While it doesn't mention Foucault (it does mention Latour!), the general critical thrust here, its

skepticism of positivism, and general debts to post-structuralism are, I think, pretty clear.

This comment is too long, so I'll stop and hope this conversation can continue at some point in the future.

Acknowledgements

Sincere thanks are due to Elli Mylonas, who edited the original text and encouraged me to post it despite my misgivings, Joan Troyano for guiding me through the process, and Natalia Cecire for her thoughtful comments and helpful suggestions as guest editor.

Originally posted by Jean Bauer on [November 3, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] This piece is unusual for me, both in style and tone. While writing it, I felt a strong need to apologize to my readers who had come to expect calmer, less combative prose. Whether this post has changed those expectations, only time will tell.

[2] October 19, 2011.

[3] These two posts were chosen as representing a particular strain of commentary. However, there were other posts written between October 19 and November 4, several of which are included in this conversation. "Who you calling untheoretical?" is also a response to the more sparsely archived Twitter stream that arose around these posts. Since the community did not use a consistent hashtag to mark their conversation, I have been unable to retrieve the relevant comments.

[4] The conversation following my blog post has shown that no offense was intended.

[5] I never define ‘theory’ in this post, a crucial failing in my rhetoric. The best definition I have since found for what I was attempting to discuss comes from a footnote in the original version of Cecire’s “[When DH Was in Vogue; or THATCamp Theory](#),” where she defines theory as “a catch-all term for thinking through the philosophical and cultural consequences of things.” Nevertheless, my failure to define my terms turned out to be a fruitful mistake, as Chris Forster used my vagueness to crystallize his own thoughts on the debate (see his insightful comments below).

[6] This point cannot be stressed enough. As an emergent strain of scholarship, digital humanities exposes the labor relations of universities while upending traditional hierarchies of knowledge production. One of the great challenges for digital humanities is learning to share credit for scholarship with a much wider range of colleagues and collaborators, regardless of official title or degrees conferred.

[7] The theoretical implications of data modeling have been discussed to great effect at the [Knowledge Organization and Data Modeling](#) Conference held at Brown University, March 14-16, 2012.

[8] [Susan Smulyan](#), Professor, American Studies Department, Brown University.

[9] I created unnecessary confusion by truncating Dr. Smulyan’s comment here for maximum rhetorical effect. This quote was never intended to suggest that databases are self-explanatory. On the contrary, I firmly believe that thoughtful, careful documentation is a fundamental requirement for digital humanities projects, because if we

do not explain the theoretical implications of our work they can easily be misinterpreted, or go unnoticed. Following her exclamation, we began a conversation on how creating and using models of reality shapes our analysis. I have written elsewhere on the theoretical implications of database design, including my essay “[Fielding History: Relational Databases and Prose](#)”.

Theory, Digital Humanities, and Noticing

When disciplines collide, as they do throughout digital humanities, the various practitioners mutually benefit from the different knowledge and skill sets that others bring to our collaborations. But there is also an inevitable gap between how different individuals working in different areas understand their various tools and techniques, and in how they understand their own thought processes. For THATCamp and digital humanities, that causes two related tensions as programmers and critics move more closely together in their joint work. The first tension is the relationship between "hack" and "yack" in the phrase often heard at THATCamp: "more hack; less yack." The second is the broader question of the role of theory in digital humanities -- its perceived absence among some and what steps can be taken to recognize what others call theory in digital humanities. Those tensions, though, will be easily resolved by addressing the disjunction between what coders know to look for as being important and what critics look for in their work to develop theory. In short, the two groups do not yet know what the other notices (in text or in code) as the first step toward understanding their work.

The prominence of hacking and building things — specifically through digital tools — led to a perception that the THATCamp workshops represent a [colonizing of the humanities by computer science](#). I would argue, however, that it is really the reverse, that the workshops are more about the humanities pushing into computer science. Not necessarily at the institutional level, but rather at the level of practices from the humanities influencing the way programmers and web developers approach their work. Many of the principles of writing code seem antithetical to much of how the humanities works. Coders like to keep things simple. We like to produce the most efficient algorithms, with very well-defined inputs and outputs. Especially with test-driven development, the code we write should have very precisely-defined functionality and purposes, and once written shouldn't change (or break).

Roughly, that's good coding, and good project management. Breaking those principles leads to [scope creep](#), and to not getting projects done. A plugin for [Omeka](#) that I started years ago is still unfinished because each time I built a new piece of functionality I noticed that it could be expanded upon to produce new, more interesting results. Parts of the input (RSS/Atom feeds) could be interpreted in a variety of ways, and so I wrote code to make the various meanings explicit. There are assumptions and ambiguities built into the input formats that I wanted to unpack and explicate for the user.

"Expand upon," "make various meanings explicit," "unpack and explicate." These are the words of a dissertation writer. I was writing code like I was writing my dissertation, which led to what a computer scientist or project manager would call scope creep. Yet I think that that kind of unpacking and examination can ultimately benefit a project. Bringing the questions that humanists tend to ask into the process of writing code can help us identify why we are writing the

code in the first place and help us recognize what promising directions or ideas are available. Or, at least, it can help us be aware of the implications of the choices we make at the level of code, and what ramifications they will have in where the final project might stand culturally. That is, thinking patterns familiar to humanists are making their way explicitly into the process of writing code and building sites and applications. That is a helpful outcome of THATCamp workshops, even though, as I found, there is a danger of it interfering with the successful completion of a project.

That's why I'm so impressed with [Jean Bauer's](#) ability to build awesome things, with theory and the humanities in mind, and actually, you know, finish them.

But I'm not terribly surprised by the reaction that she describes in "[Who You Calling Untheoretical?](#)" of people asking, "But where's the theory?" I'm not surprised, because, in general, I don't think theorists know what to notice yet. Our graduate programs are naturally still dominated by fairly traditional "texts," and theory rooted in them. And our theory has done a good job of complicating various notions of "text" (note to non-humanist coders: complicating ideas is a good and productive thing to theorists). Whatever our theoretical approach to whatever we are calling a text, we know what to notice in them. That's what much of our grad school training in theory is about -- what do we need to notice in the text, and how does our theory help us describe what we have noticed?

This is not surprising either, because unlike any of the various theoretical notions of text floating around, most people don't have an experiential grounding in database as text or API as text. One doesn't need to be a poet to do literary criticism on poetry, though the poet and critic will share a common knowledge base. But (unless you are a medievalist or classicist or doing comparative literature) you don't

need any special training to be able to begin noticing things in the text. There is a base line of experience with reading and discerning meaning that you can start with.

Imagine farther back to an introductory course in literature or in writing. These courses have a goal of teaching students to be more aware of the texts that they read and write. That training often starts with skills at close reading, which works well because we can guide students to noticing patterns in the text by starting with what is already accessible to them: the basic meanings of words they already know and use (but usually in an unreflective way). Noticing can start at the easily accessible moment of observing a word's meaning. We can start training up our reading skills because we don't need to be taught the basic meanings first. From there, we bring in complications in an effort to show that the first, most accessible basic meaning is not the end of the story. We encounter the question, "Why can't I just read for fun?" exactly when students begin to move beyond that easiest, immediately accessible meaning and begin to make what they first perceive as the obvious meaning more complicated. But it starts with helping them to notice the meaning they immediately saw, without the additional training.

Of course, different techniques of noticing become more important as we continue our training. Developing skills of noticing in Toni Morrison and in *Beowulf* require building upon more sophisticated techniques of noticing. As we train up, the required knowledge and skills diverge by discipline and subdiscipline until many of our core skills are mutually inaccessible.

I see a very different trajectory in the digital world. A good user interface is designed specifically so that you don't have to deal with the inner workings of the application. In general, people should not see the internal structures of an application — the database, the public and

private methods in the core code. Unlike our, or our students', first experiences with reading in more interesting and complicated ways, the first starting point — the language — is specifically (and, in closed-source applications, legally) hidden away from us. And so, there is no ability to even begin noticing what's notable.

This is where crit-code studies seems promising. There might be strong analogies to crit-code training and methods and comparative literature, but that's something for someone more knowledgeable about both of them to figure out.

As digital humanities begins to tackle what theory is and does in our discipline, we are in the interesting position of disciplinary *convergence*, rather than the divergence that typically characterizes our training. In broad strokes, training in the humanities begins with noticing the immediate meaning of a word or phrase in our early undergraduate courses, then, through graduate school and publications, learning how to unpack and explicate meaning with increasingly complex methodologies, each of which depend on different skills and techniques of noticing.

With the collision of the worlds of computer science and humanities in current digital humanities, we have a situation that calls for a confluence of noticing skills from different specialization that are in new kinds of conversations. To computer scientists, code will be accessible in a way that it is not to humanists without familiarity with programming languages. They will have a basic ability to start noticing things analogous to the basic ability to start noticing that our intro-level humanities students have. Instead of reading it for fun, they will want to read it for efficiency or for good application design. The humanists unfamiliar with code won't have that opening ability to start noticing, but once they do, they will have important things to say about

what is implicit in the code and what needs to be explicated, and not just for fun or efficiency.

I'm not sure I'd go so far as to say that to do theory in/on digital humanities one needs to learn to code or design a database. But one does need some training to be able to start noticing the difference between two data models that at surface appear to describe the same things. And, coders should be ready to learn what useful things theorists can offer that, despite a first appearance of scope creep, might just be valuable things to consider building into the code.

As a concrete example, I sometimes fret over the centrality of the item in [Omeka](#). An "item" is the fundamental unit of information, and we have lots of ways to describe them with metadata, mostly Dublin Core. But "items" can become complicated very quickly. Some of my fellow colleagues at the [Roy Rosenzweig Center for History and New Media](#) are struggling with how to put scrapbooks into Omeka. Is the scrapbook the item, or is each individual thing in the scrapbook an item? Where do pages in the scrapbook fit in? I'm guessing that there is theory that could be useful here, and could lead to an Omeka plugin that is designed to implement a theoretical approach to scrapbooks that lets us work with a more complex notion of "item" in our data model. I need some theory to help me notice things about scrapbooks, and to help me notice things about Omeka's notion of "item."

Producing a plugin that complicates Omeka's model of "item" by consciously building theory into the code would be a great code hack in harmony with theory yack.

Acknowledgements

[The original blog post](#) that developed into this piece was primarily a response to Natalia Cecire's "[When DH Was In Vogue; or, THATCamp](#)

[Theory](#)", Jean Bauer's "[Who You Calling Untheoretical?](#)", and the resulting comments and twitter conversations.

Many thanks to PressForward, and especially to Natalia Cecire, for their help in developing this from the original post into its current form.

Originally published by Patrick Murray-John on [November 10, 2011](#).
Revised for the *Journal of Digital Humanities* March 2012.

Digital Humanities as Thunderdome

Recently at a workshop on digital tools for the humanities, a Stanford graduate student rather poignantly noted that oftentimes collaboration with computer scientists felt more like colonization by computer scientists. This statement, even if not true, is far too sharp to ignore. Frankly, I think it is true. Not long after that workshop, I attended a [THATCamp](#), where I spent my time teaching folks how to use [Gephi](#), and I tried to spend some time telling them that the network they create is the result of an interpretive act. I don't think they cared, I think they just wanted to know how to make node sizes change dynamically in tandem with partition filters. This is an issue that has concerned me for some time: the way wholesale importation of digital tools, techniques and objects into humanities scholarship tends to foster a situation where rich, sophisticated problems are contracted to fit conveniently into software.

I love Gephi; that's obvious, but it isn't built for humanists, because *nothing is truly built for humanists*; the closest we can get is something built *by* humanists. At their core, the systems, protocols and logical framework that are our digital world are created by and for a very pragmatic minority of our society. Engineers and scientists, by

and large, do not problematize the "best practices" developed over a long and successful period of creating digital tools and objects for design, medical care, advertising, manufacturing, and so on. Yet the very languages, standards and applications that are used for digital humanities scholarship are a result of various collaborations between engineers and their professional clients. Add to this the fact that there is little in the way of domain specialization in the field of humanities scholarship among software engineers, and you end up with a situation I once described as trying to use a satellite built for mapping elevation to instead map culture.

Gephi is not built by humanities scholars, nor is it built for humanities scholars, and as such it has core logic that requires subversion in order to represent complex and uncertain digital humanities arguments. I point to Gephi because I use it all the time, and I know the people who code it, and I've written some code to extend it, but this is a fundamental fact not only of all software that wasn't written specifically for a scholarly humanities audience, but for all software itself, which is still embedded in the pragmatic, engineering mindset from which it was born. Johanna Drucker suggests that scholarly objects born of this process lack "many humanities principles developed in hard-fought critical battles of the last decades," offering her "short list":

the subjectivity of interpretation, theoretical conceptions of texts as events (not things), cross-cultural perspectives that reveal the ideological workings of power, recognition of the fundamentally social nature of knowledge production, an intersubjective, mediated model of knowledge as something constituted, not just transmitted. For too long, the digital humanities, the advanced research arm of humanistic scholarly dialogue with computational methods, has taken its rules and cues from digital exigencies.[1]

There *are* groups who have more experience with adapting digital tools and objects to their work than digital humanists, though I don't think they are the answer. Archaeology, for example, is traditionally pretty pragmatic in its use of technology, and museums and business are oriented toward a different audience than humanities scholarship at a Research 1 university. The digital humanities "segment of the market" doesn't, as yet, have a corresponding set of domain specialists in computer science to help fashion UI/UX, data modeling and requirements for proper digital humanities scholarship and I really don't think we ever will—there simply isn't the market for it; the work is too sophisticated, specialized and absurd. The best we can do, if we go the traditional route, is to latch on to best practices from journalism, or public humanities like museums and library science. All of these, I think, carry with them the problems of authorial bias toward simplifying scholarly humanities problems.

The other option is to not touch the filthy digital, which would keep humanists clean but make them fundamentally divorced from the modern world. A third path is for humanities scholars themselves to pick up coding, write weird and not-at-all pragmatic software and, perhaps, create standards through practice or, more likely, just create lots and lots and lots of weird code that better describes queer black artists in the twenties or a republic of letters or Walt Whitman or James Joyce or Søren Kierkegaard.

Regardless, the first step is awareness of what a tool or method is doing and how it will inflect your research. I'm concerned that humanities scholars show a willingness to defer to tools, but I'm more concerned that they may simply surrender to tool builders. When I got to Stanford, I felt superior to long-tenured faculty members because I knew how to code and they didn't, and this circumstance was reinforced by the fact that they had to ask me whether something was

possible. That's a horrible burden to put on a young scholar or alt-ac type like me. It's quite the temptation to answer questions like those as if I really knew all the possibilities of digital representation of humanistic inquiry. Because, really, the answer I'd give is only based on my limited coding skills and my even-more-limited understanding of the domain of the scholar I'm supporting.

What makes this doubly dangerous is that the sense of power and authority afforded in this situation is a useful tonic for the lack of official respect accorded to alt-ac staff members who work with faculty in Research 1 universities. I've tried, I hope with some success, to actively combat it by engaging my faculty collaborators in the nitty-gritty details of the logical systems that are being put in place to translate their work into the digital realm, all the while foregrounding the fact that, like any act of translation, it is interpretive and limited. I now playfully mock and goad humanities scholars when they claim to be incapable of understanding models and code, because I want to put an end to this dance being done to show some level of respect for, but also a willingness not to intrude on as well as a separation from, the domain of the tech support.

As my work has involved ever more input from humanities scholars on the most fundamental functions of the models and interfaces that they create, I have become far more aware of the years of work that scholars have put into truly understanding their fields. My knowledge of how to code does not necessarily overcome my relative inexperience in the understanding of or engagement with humanities scholarship. However, this isn't meant to be some kind of self-abasing paean to the great and glorious tenured humanities faculty – if they want to do sophisticated digital humanities work they'll need to learn how code works, if not actively become coders.

As it stands, it is all too common that a respected scholar with years of experience in their subject matter is constrained by a masters student in computer science who, in the best of cases, is a charitable ally with little knowledge of the domain and its many, thorny issues.[2] This situation is much like that found in the classic film [Mad Max: Beyond Thunderdome](#). For those not familiar with this particular artistic work, its title comes from an arena where “two men enter, one man leaves.” I think we’re replaying that moment over and over again in the digital humanities, with pragmatic, clean, idealized, “best practices” on one side and the queer, messy, uncertain and post-modern on the other. It’s not nearly as good an analogy as [Cecire’s Harlem Renaissance](#), I know, but it has the benefit of being available via BitTorrent and YouTube. It isn't really a gladiatorial death match between computer scientists and humanities scholars, though, because the two combatants are much more primal than that--a *poesis* (poetic, emergent, and contingent) form of knowledge expression to struggle with the *techne* (technical, crafted, or objective). It’s about transitioning the representation of humanities knowledge out of text and into the digital without transforming it into a simplified version.

Success is a pragmatic ideal, but in this case I'm willing to employ it in an attempt to transition from defining the digital humanities into defining success for the digital humanities. I see it as creating a humanities that is not as simplistic and flat and technical as that envisioned by engineers and computer scientists, but as rich and sophisticated and poetic as that described in our libraries and seminar rooms and long discourses. The liminal quality of the narrative text format that we've used to present humanities knowledge is something to foster and integrate into a successful scholarly digital object, and not something that needs to be stamped out because it does not fit into the patterns of data management and manipulations established by the early adopters of digital tools, objects and methods. And I need to be

clear, it's not something we should maintain because it is a cultural artifact, but because it allows for a more accurate, if less precise, representation of human experience.

Originally published by Elijah Meeks on [November 5, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] Johanna Drucker, “[Blind Spots](#),” *The Chronicle of Higher Education* (April 3, 2009): B6.

[2] I think Molly Des Jardin [provides a telling perspective on the view of the humanities among early-career computer scientists](#).

Words and Code



@foundhistory

Tom Scheinfeldt

DH arguments are encoded in code. I disagree with the notion that those arguments must be translated / re-encoded in text.

10 Nov via Seismic [★ Unfavorite](#) [↻ Retweet](#) [↩ Reply](#)



@rybesh

Ryan Shaw

@foundhistory @ncecire If you can't explain to me in words how your code works, you don't really know how it works.

11 Nov via Twitter for Mac [☆ Favorite](#) [↻ Retweet](#) [↩ Reply](#)

Please Write it Down: Design and Research in Digital Humanities

[Tom Scheinfeldt provocatively suggested](#) that "DH arguments are encoded in code" and that he disagrees "with the notion that those arguments must be translated / re-encoded in text." I don't think this is how this works.

What I see as the key issue is not so much whether digital humanists need to "re-encode" their work in writing. Digital humanists, like reflective designers of all stripes, are already doing a lot of writing. They are creating documentation, making wireframes, etc. The question here is: what kinds of writing should humanities scholars who design software and make things in code be doing?

Writing is Thinking, and Designers Write Things Too

Everybody working on a digital humanities project needs to be writing. I am suggesting that this is simply a fact of life. If you don't have at least [a one-pager for your project](#), then you don't have a project; you are just fiddling around. In fact, purposeful design necessitates the creation of documentation at nearly every step. As I recently suggested, [every document and artifact that you create in the process](#)

[of design could serve as a new genre of humanities scholarship](#). For starters, practically everything in Dan Brown's [Communicating Design](#) already looks like the kinds of things we already write.

As I see it, it is not that you need to translate what you did in code into text. Instead, to have made something interesting in code you probably went through a reflective process that inevitably created a wake of valuable texts that were central to both the creation of the argument the code made, and potentially the most viable communication of that argument. You probably only need to clean them up a little bit. Even better, many projects are the result of grant-funded work. In those cases, the text already exists since the creator needed to explain what the thing they were going to make was supposed to do.

With this said, I would also suggest that at the end of a project ([or whatever it is we are calling done](#)), taking time to sit down and write out what you learned is an invaluable reflective practice. In my own experience, this is far from being the moment when you translate something you already knew into another format; rather this is the moment that crystallizes what you actually learned. This is not about writing it up. Taking a few moments at the end of a project to reflect on what you wanted to accomplish, what actually happened, and what you learned from the process is critical not only for communicating results, but for really coming to know them.

With Design and Humanities Research We are Still Only Beginning

So people who make stuff have to write a lot about what they are doing as part of the process of making stuff. This kind of writing is simply part of being a reflective designer. But I think we are only scratching the surface of how purposefully thinking about the process of design could become a key part of humanities scholarship.

I have my feet in both the digital humanities world and the world of educational research, so I would like to point digital humanists to an ongoing conversation in instructional technology about design and research practices. About twelve years ago, educational technologists started talking about something they call design-based research. The idea is that instead of contriving wonky experimental designs, it would be better for researchers to adopt the role of designer and think through how formalizing the iterative practice of design could serve as a basis for research methods. The idea behind design-based research is that there is some kind of hybrid form of doing, theorizing, building and iterating that we should turn into a methodology.

Two articles summarize this conversation nicely: [Design-Based Research: An Emerging Paradigm for Educational Inquiry](#) (PDF) from The Design-Based Research Collective and published in *Educational Researcher* in 2003, and [Design-Based Research: Putting a Stake in the Ground](#) (PDF) by Kurt Squire and Sasha Barab, published in *The Journal of the Learning Sciences* in 2004. The Design-Based Research Collective's piece suggests how theory, practice and method coalesce in research-based design.

Design-based researchers' innovations embody specific theoretical claims about teaching and learning, and help us understand the relationships among educational theory, designed artifact, and practice. Design is central in efforts to foster learning, create usable knowledge, and advance theories of learning and teaching in complex settings.

In short, yes; designs always have explicit and implicit arguments inside them. However, reflective designers produce a range of artifacts and documents during the process of design that, if shared, could both help them become better designers, and help others learn to become better designers. Further, the concept of design-based research pushes us to think more deeply, and not simply absorb the design practices of

others. What might a design-based research method look like if we translated it from the educational context and into the context of a particular humanities research question?

Originally published by Trevor Owens on [November 11, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Building and Sharing (When You're Supposed to be Teaching)

Much of what I do in my classroom doesn't necessary count as "digital humanities." I certainly don't present my classes as digital humanities classes to my students---or to my colleagues, for that matter.

If anything, I simply say that we'll be doing things in our classes they've never done before in college, let alone a literature class. And literature is mostly what I teach. Granted, I teach literature that lends itself to digital work: electronic literature, postmodern fiction, and even videogames. We do a great deal of close readings in these classes. It's familiar, even comfortable territory for my students. But we also---and this is what surprises my students---spend much of our time building and sharing.

In fact, if I were to change the title of this essay to reflect my students' perspectives, it might look something like this:

Building and Sharing (When We're Supposed to be Writing)

And at the end of this title would come one of the greatest unspoken assumptions both students and faculty make regarding writing: writing:

((For an Audience of One))

So the "sharing" part of my title comes from my ongoing effort---not always successful---to extend my students' sense of audience. I'll give some examples of this sharing shortly, but first I want to address the initial word of my title: Building. Those who know me are probably surprised that I'm emphasizing "building" as a way to integrate the digital humanities in the classroom. One of the most popular pieces I've written in the past year is [a blog post](#) decrying the hack versus yack split that routinely crops in debates about the definition of digital humanities. In this post, I argued that the various divides in the digital humanities, which often arise from institutional contexts and professional demands generally beyond our control, are a distracting sideshow to the true power of the digital humanities, which has nothing to do with production of either tools or research. *The heart of the digital humanities is not the production of knowledge. It's the reproduction of knowledge.*

The promise of the digital is not in the way it allows us to ask new questions because of digital tools or because of new methodologies made possible by those tools. The promise is in the way the digital reshapes the representation, sharing, and discussion of knowledge.

And I truly believe that this transformative power of the digital humanities belongs in the classroom. *Classrooms were made for sharing.* Where, then, does the "building" part of my pedagogy come up? How can I suddenly turn around and claim that building is important when I had previously argued the opposite, in a blog post that has shown up on the syllabuses of at least [six introduction to the digital humanities courses](#)?

I need to explain what I mean by building. Building, for me, means to work. And when I say work, I mean the opposite of thinking. I get this

idea from a short essay by Peter Sallibrass that appeared in the [PMLA in 2007](#). Sallibrass's article has the provocative title "Against Thinking," and in it, he argues that we think too much and don't work enough.

Thinking, according to Sallibrass, is the hobgoblin of big minds. Thinking is boring, repetitious, and "indolent" (1583). On the other hand, working is "easy, exciting," and "a process of discovery" (1583). Working is challenging.

This distinction between thinking and working informs Sallibrass's undergraduate pedagogy, for example, the way he trains his students to work with archival materials and the English Short Title Catalog. In Sallibrass's mind, students---and in fact, all scholars---need to do less thinking and more working. "When you're thinking," Sallibrass writes, "you're usually staring at a blank sheet of paper or a blank screen, hoping that something will emerge from your head and magically fill that space. Even if something 'comes to you,' there's no reason to believe that it is of interest, however painful the process has been" (1584). This is a key insight that students and scholars alike need to be reminded of: tortured and laborious thinking does not automatically translate into anything of importance.

Sallibrass goes on to say that "the cure for the disease called thinking is work" (1584). In Sallibrass's field of Renaissance and Early Modern literature, much of that work has to do with textual studies, discovering variants, paying attention to the material form of the book, and so on. In my own teaching, I've attempted to replace thinking with building---sometimes with words, sometimes without.

I'll share a few examples here from my own teaching, which broadly fall into two categories: *collaborative construction* and *creative analysis*. By *collaborative construction*, I mean a collective effort to

build something new, in which each student's contribution works in dialogue with every other student's contribution. A key point of collaborative construction is that the students are not merely making something for themselves or for their professor. They are making it for each other, and, in the best scenarios, for the outside world. Collaborative construction obliterates that insular sense of audience inherent in more conventional student assignments. As for the concept of *creative analysis*, I mean that as a kind of antidote to the vacuous and shape-shifting term "critical thinking." Creative analysis is the practice of discovering knowledge through the act of creation---through the making of something new. Rather than having students write papers, which often involves the worst aspects of thinking that Sallibrass derides, I ask the students to do something they find severely discomfiting: creating something new for which no models exist.

As examples of collaborative construction, I offer up my students' [Portal Exhibit](#) and a cross-campus effort to [renetwork of House of Leaves](#). With the *Portal* Exhibit, students in my George Mason University Honors College course on [Technology in the Contemporary World](#) used [Omeka](#) to build an online exhibit dedicated to the groundbreaking game *Portal*. The exhibit was entirely student-designed, and though the results fell short of my initial vision for the exhibit, the students encountered a number of practical and epistemological challenges that deepened their understanding of the both the game itself and the way we talk about and make sense of videogames more generally.

A more decentralized version of constructive collaboration occurred in my Fall 2011 [Post-Print Fiction](#) class, in which my students read Mark Z. Danielewski's *House of Leaves* alongside four other classes at four universities or colleges (Converse College, Temple University, Emory

University, and the University of Mary Washington). All five classes then participated in an online forum that strove to replicate as closely as possible the original online *House of Leaves* discussion forum, which at its peak had hundreds of participants and thousands of posts. Our classes were, in a sense, rebooting the forum.

As examples of creative analysis, I want to point to several types of mapping and modeling projects I've used. In a postmodern fiction class, I've had students build [abstract models of a novel](#) (obviously inspired by Franco Moretti's notion of distant reading). In a videogame studies class I've likewise asked students to design [an abstract representation](#) of an NES game, a kind of model that would capture some of the game's complexity and reveal underlying patterns to the way actions, space, and time unfold in the game. As I've [reflected upon](#) elsewhere, I try with such projects to turn my students into aspiring Rauschenbergs, "assembling mixed media combines, all the while through their engagement with seemingly incongruous materials, developing a critical thinking practice about the process and the product." In the videogame class I'm also experimenting with [game design](#) projects as alternatives to traditional final papers. The very act of designing a game instead of writing a final paper changes the students' sense entirely of what they're doing and who their audience will be. Students know that a final paper will be read---hopefully---by only one person (if that). A game, however, already presumes an audience.

If I were to say what unites these various forms of building in my classroom, I might use the term "deformance," a portmanteau [coined in 1999](#) by Lisa Samuels and Jerry McGann. A combination of "performance" and "deform," deformance is an interpretative concept premised upon deliberately misreading a text, say, reading a poem backwards line-by-line. More recently, Stephen Ramsay demonstrates

in [Reading Machines](#) how computers allow scholars to practice deformance quite easily. I would add (and I doubt Ramsay would disagree) that it's not only texts that can be deformatively reshaped, nor are computers necessary tools for deformance. As my students build---both collaboratively and creatively---they are also reshaping, and that very reshaping is an interpretative process. It is not writing, or at least not only writing. And it is certainly not only thinking. It is work, it has an audience, and it is something my students never expected.

Originally published by Mark Sample on [October 19, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Originally a lightning talk given on October 18, 2012 as part of CUNY's [Digital Humanities Initiative](#).

Marked Bodies, Transformative Scholarship, and the Question of Theory in Digital Humanities

In October 2011, Natalia Cecire’s [off-the-cuff suggestion of a THATCamp Theory](#) set off a ferment of planning and arguing in the digital humanities community. It sounded like a great idea to me. Beginning with [a session on “diversity in digital humanities”](#) at THATCamp SoCal in January 2011—well attended both in person and remotely—I had been collaborating with an amorphous group of scholars engaged with critical cultural studies, queer studies, and ethnic studies in the context of the digital. We had been thinking about ways to connect the ethic of making that is central to digital humanities with a greater self-consciousness about the way everything is structured and its cultural politics; I was keen to continue and broaden that conversation.

Yet the discussion that emerged from Cecire’s post turned out not, by and large, to be about theorising the work of the digital humanities in this sense. I was unsettled by some of the ways “theory” came to traffic in the conversation: both by the defensive, sometimes even accusatory, tone in which the term was uttered, and by the histories of exclusionary practices it was held to evoke.

Ted Underwood’s post “[On transitive and intransitive uses of the verb ‘theorize,’](#)” for example, described how the demand for ‘theory’ can be used as a demand for control:

a tenured or tenure-track faculty member will give a talk or write a blog post about the digital humanities, saying essentially “you’ve got some great tools there, but before they can really matter, their social implications need to be theorized more self-consciously.” Said professor is then surprised when the librarians, or academic professionals, or grad students, who have in many cases designed and built those tools reply with a wry look.[1]

The reason for this, as [Miriam Posner](#) recently tweeted, is that “[theory has been the province of scholars,](#)”[2] while “[the work of DH has been done by staff.](#)”[3] So when you say “those tools need to be theorized,” you are in effect saying “those tools need to be appropriated or regulated by someone like me.”

Underwood places this “vague, intransitive” call for practices “to be theorized” in opposition to the way that digital humanities operates “an insurgent challenge to academic hierarchy, organized and led by people who often hold staff positions.” Jean Bauer similarly insisted, in her provocatively titled post “[Who You Calling Untheoretical?](#)” that the architects of digital projects are often fully aware of their theoretical implications. She writes that to make digital scholarly work is to make theory—of a kind that cannot be separated from its material context: the kind that Underwood would call transitive.

Underwood goes on to write that the difference digital methods make to the practice of humanities scholarship will require some intransitive theorizing. But the question of theory in these posts is always a question of academic recognition; even the “insurgents” are firmly located as laborers within the university. And, even within the critique of intransitivity, “theory” seemed to be operating without much specific

content. What about the kinds of theory that link up to activist projects, that unpack the politics of academic knowledge production itself and the relationship of its hierarchies to cultural, social, economic difference?

In a [summary of the various discussions](#), Roger Whitson complained that “[b]eing ‘theoretical’ or dealing with ‘theory’ can sometimes be conflated with revolution, sex, and power without actually being any of those things.”^[4] But what I saw, even in discussions that aired vital critiques, was the invisibility of how any of those things might be linked to discussions of the less glamorous matters of race, class, and gender—concerns that are emphatically not only the domain of tenured and tenure track professors, nor even only of academic faculty, students, and staff.

Cecire’s post linked to a set of conversations in which the question of theory was intimately involved with these concerns, though in the subsequent conversation it seemed largely to disappear from view. She discussed Micha Cárdenas’s provocative “[Digital Humanities: Hot Sellable Commodity or Place of Counter-Hegemonic Critique?](#),” a response to the Los Angeles Queer Studies conference and to a panel on digital theory and praxis in which Cárdenas and I participated along with Margaret Rhee and Amanda Phillips. The question in Micha’s post was not ‘where is the theory in digital humanities projects?’ As a scholar, artist, and digital practitioner, Cárdenas takes Underwood and Bauer’s insights as a starting point; she wants to know not where the theory is, but what the theory does. She asks about the status of the digital humanities, theory and praxis alike:

Do you think there is often something very conservative, even sellable, that is appealing to corporations or to university regents or investors, that is often present in discussions of the digital humanities? Do you think there is still some radical potential for queer theory or new media or the

digital humanities to disturb hegemonic systems of power that facilitate violence against certain groups of people every day and protect the interests of others?^[5]

These are also the questions we were asking in our diversity session at THATCamp SoCal. They are the questions that theorists, scholars, and practitioners, including Anna Everett, Lisa Nakamura, and Tara McPherson, have been asking for years. They are questions that stubbornly refuse to appear at the center of the bodies of knowledge and practice, the conversations that shape what we know as digital humanities. And they are the questions around which I, together with Amanda Phillips, Tanner Higgin, Marta S. Rivera Monclova, Melanie E. S. Kohnen, and Anne Cong-Huyen, organized [a panel chaired by Anna Everett](#) for the American Studies Association (ASA) conference in October 2011. Cecire quoted the first paragraph of our description:

In an era of widespread budget cuts at universities across the United States, scholars in the digital humanities are gaining recognition in the institution through significant grants, awards, new departments and cluster hires. At the same time, ethnic studies departments are losing ground, facing deep cuts and even disbandment. Though the apparent rise of one and retrenchment of the other may be the result of anti-affirmative action, post-racial, and neoliberal rhetoric of recent decades and not related to any effect of one field on the other, digital humanities discussions do often elide the difficult and complex work of talking about racial, gendered, and economic materialities, which are at the forefront of ethnic and gender studies. Suddenly, the (raceless, sexless, genderless) technological seems the only aspect of the humanities that has a viable future.^[6]

Our ASA panel insisted that the future of the technological humanities will never be a raceless, sexless, genderless, or apolitical one. It brought together emergent and established scholars working on and with technology in order to do scholarly work that aims to support (if not actually foment) social and cultural transformations that might, in

Cárdenas's words, "disturb hegemonic systems of power that facilitate violence against certain groups of people every day and protect the interests of others." Our panel was titled "Transformative Mediations" in reference to this; we used #transformDH as a hashtag to document it. In the few short months since ASA, #transformDH has solidified as a collective with several writing projects in the making, a descriptive term for digital humanities projects with a critical cultural studies orientation, and something of a rallying cry. And, to come to the point of this piece and this collection, #transformDH is theory.

At ASA, the #transformDH collective-in-the-making was demanding theory for the digital and for the humanities, but we were not using the term intransitively. We were talking about queer, trans, butch, femme, critical race, women of color, Asian American, Puerto Rican theory. With a slightly different group of scholars in the room, those adjectives would have changed, but their tangibility would remain. We were talking about marked bodies, systemic social hierarchies, and transformations in a very specific and material sense. This was not 'Theory' as a vague revolutionary concept all too easily written off by the image of turtlenecked graduate students sitting around talking about Foucault that it conjures. We were talking about theory as making, about making objects that critique, that *are* critique, that are transformative reimaginings of the world.

For an example, we might look to [Cárdenas's artwork](#), which includes wearable electronics figured as devices that would enhance the safety of sex workers by giving them access to support networks not mediated by the state.[7] One of the most important parts of this kind of theory, to me and to many members of the #transformDH collective, is that it is not only made in the academy. What conversations, artforms, databases, and archives might do the work of a transformative digital

humanities, though they lack the institutional status to be named as such?

When I looked at the discussions about theory and digital humanities that emerged around the birth of [THATCamp Theory](#), I found myself faced with my cohort's disappearance. Where did we go? Where did our marked bodies—our politics and our specificity—go? I wondered whether we might need a term different from "theory" in order to become visible. A [tweet by Jentery Sayers](#) suggested as much, as he drew attention both to his own work in creating the social justice focus of THATCamp Pacific Northwest and to Alan Liu's important interventions in his MLA 2011 presentation "[Where is Cultural Criticism in the Digital Humanities?](#)" and the [4Humanities](#) project it launched.[8]

But I remain attached to the term "theory" and to the possibility that it can be democratized. I want all these forms of critical making and the analysis that accompanies them to be part of the "theory" conversation, if there is a "theory" conversation to be had. I fear that their specificities may be dismissed as irrelevant identity politics, and I want to insist that they not be. The markedness of our bodies (even, perhaps especially, those who might experience their bodies as unmarked) is not a marginal or irrelevant concern. This is the heart of things, the center from which our digital work radiates. And these concerns are not exclusive to the digital. Embodied theorizing is especially visible in the zones where scholarship and practice overlap—art, performance—but we never leave our bodies and their cultural mattering behind.

Part of the conversation about how we make theory has to be a conversation about which forms of theory-rich making are recognized and institutionally supported and which are not; about whether there are clear cut lines between digital humanities scholarship, digital

media art, and digital media everyday practice, other than the question of where the funding comes from. This brings us back to the questions of theory and power that Underwood, Bauer, and Miriam Posner (as cited by Underwood) have raised. There are unstated hierarchies of labor when we differentiate between who does the work of making versus who conceptualizes or “theorizes” a project.

There are even more hierarchies involved when we think about what counts as a “project” deserving of labor other than basic conceptualization. Paying attention to race and to the bodies that do theory cast this into sharp relief. One of #transformDH’s instigators, [Marta S. Rivera Monclova](#), has struggled in making the necessary theory for her planned project on multilingual Puerto Rican poetry visible. Her project is concerned not only with translation but with the transformations that multilingual racialized and gendered subjectivity engage and produce; to develop a digital humanities project that can express this, she is crafting transformative theory.

In the end, for me to insist that it be possible to mean #transformDH when we say “theory” is a strategic intervention, of course. Especially for those of us who have passed through graduate school in the humanities, theory can operate in a multitude of ways, producing exclusions and doing violence as often as it gives voice to the excluded and offers ways of recognizing previously unnoticed histories of oppression. The conversation about THATCamp Theory sprouted some beautiful metaphors for this. In a [conversation](#) between Cecire and Posner that Whitson collated, theory figured as something that could be “wielded,” like a weapon, terms brought forth to silence those without the cultural capital to use them. Yet it could also be held “softly, like a bunny,” put in the hands of those who will gain much by its tools. The liveliness of the bunny metaphor grew larger with Cecire’s tongue-in-cheek suggestion that theory might really be the Loch Ness

monster: lurking under the surface of everything, ready to bite, yet also something we constantly look for but never find.

The conception of theory I have been arguing for here, which comes both from the academic realms that have nurtured #transformDH and from a range of nonacademic institutions and locations, mixes all these metaphors. I want to think about the digital, the humanities, and the digital humanities with the help of an [awkwardly handcrafted pet theory monster](#), one that I may wield from time to time, but only if I nurture it and encourage it to play well with others. Yet even as I don’t want to eclipse, erase, or eat up other kinds of theories, I hope that our #transformDH theory monster might end up being more efficient and dangerous than she looks. Nessie does, after all, have teeth.

Originally published by Alexis Lothian on [November 4, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

Notes:

[1] Ted Underwood. “On Transitive and Intransitive Uses of the Verb ‘theorize.’” *The Stone and the Shell*, October 23, 2011. <http://tedunderwood.wordpress.com/2011/10/22/on-transitive-and-intransitive-uses-of-the-verb-to-theorize/>.

[2] Miriam Posner. “I guess what nags at me abt the critique of DH coming from #2011ASA is that “theory” has long been the province of scholars, while [+].” October 22, 2011. <https://twitter.com/#!/miriamkp/status/127765652458053632>.

[3] Miriam Posner. “the work of DH has been done by staff. And there is a hierarchy. I worry that staff will be edged out of the

conversation. [-] #2011ASA" October 22, 2011. <https://twitter.com/#!/miriamkp/status/127766100795596802>.

[4] Roger T Whitson. "THATCamp Theory Bunnies." Roger T. Whitson, PhD, November 4, 2011. <http://www.rogerwhitson.net/?p=1230>.

[5] Micha Cárdenas. "Is the Digital Humanities a Hot, Sellable Commodity? Or a Place for Counter Hegemonic Critique?" HASTAC, October 18, 2011. <http://hastac.org/blogs/michacardenas/2011/10/18/digital-humanities-hot-sellable-commodity-or-place-counter-hegemonic->

[6] Alexis Lothian. "Conference Thoughts: Queer Studies and the Digital Humanities." Queer Geek Theory, October 18, 2011. <http://www.queergeektheory.org/2011/10/conference-thoughts-queer-studies-and-the-digital-humanities/>.

[7] Micha Cárdenas. "Wearable Electronics as Femme Disturbance." UCLA, 2011.

[8] Jentery Sayers. "@alothian Indeed, Which Is Why I'm Inclined to Echo Liu & Say 'Cultural Criticism.' Every Field Has Theory? Lots to Unpack, for Sure.", November 3, 2011. <https://twitter.com/#!/jenterysayers/status/132196527862059008>.; Alan Liu. "Where Is Cultural Criticism in the Digital Humanities" presented at the Modern Language Association, Los Angeles, CA, January 7, 2011. <http://liu.english.ucsb.edu/where-is-cultural-criticism-in-the-digital-humanities/>.

TIM SHERRATT

Where Are the Philosophers? Thoughts from THATCamp Pedagogy

In 1901, one of the first acts of the Commonwealth of Australia was to create a system of exclusion and control designed to keep the newly-formed nation 'white'. But White Australia was always a myth. As well as the Indigenous population, there were already many thousands of people classified as 'non-white' living in Australia — most were Chinese, but there were also Japanese, Indians, Syrians and Indonesians.

Here are some of them...



["The Real Faces of White Australia"](http://www.bineaustralia.org/faces)

The administration of what became known as the White Australia Policy created a huge volume of records, much of which is still preserved within the [National Archives of Australia](https://www.naa.gov.au/). These photographs are attached to certificates that non-white residents needed to get back into the country if they decided to travel overseas. There are thousands upon thousands of these certificates in the Archives. Thousands of certificates representing thousands of lives — all monitored and controlled.

But it is too easy to see these people as the powerless victims of a repressive system. There were many acts of resistance. Some argued against the need to be identified 'just like a criminal'. Others exercised control over their representation, submitting formal studio portraits instead of mug shots.

Mr and Mrs William
Yinson Lee

National Archives of Australia:
A1, 1916/21599



Most commonly and most powerfully, people resisted the policy simply by going ahead and living rich and productive lives.

My partner, [Kate Bagnall](#), is helping to rewrite Australian-Chinese history by overthrowing the stereotype of the culturally isolated Chinese man living a lonely, meagre existence surrounded by gambling and opium dens. By mining the available records, by reading against

the grain of contemporary reports, and by working with family historians, Kate is documenting their intimate lives—their wives, their lovers, their families and descendants—the sorts of relationships that sent a shudder through the edifice of White Australia. Power can be reclaimed in many subtle and subversive ways.

‘The real face of White Australia’ [is an experiment](#). It uses facial detection technology to find and extract the photographs from digital copies of the original certificates made available through the National Archives of Australia’s collection database. The photographs you see here come from just one series, ST84/1. There’s no API to the collection so I reverse-engineered the web interface to create a script that would harvest the item metadata and download copies of all the digitised images. There are 2,756 files in this series. On the day I harvested the metadata, 347 of those files had been digitised, comprising 12,502 images. It took a few hours, but I just ran my script and soon I had a copy of all of this in my local database.

Then came the exciting part. Using a facial detection script I found through Google and an open source computer vision library, I started experimenting with ways of extracting the photos. After a few tweaks I had something that worked pretty well, so I pointed my aging laptop at the 12,502 images and watched anxiously as the CPU temperature rose and rose. It took a few emergency cooling measures, but the laptop survived and I had a folder containing 11,170 cropped images. About a third of these weren’t actually faces, but it was easy to manually remove the false positives, leaving 7,247 photos.

These photos. These people.

With my database fully primed and loaded it was just a matter of creating a simple web interface using Django for the backend and Isotope (a jQuery plugin) at the front. Both are open source projects.

All together, from idea to interface, it took a bit more than a weekend to create, and most of that was waiting for the harvesting and facial detection scripts to complete. It would be silly to say it was easy, but I would say that it wasn’t hard.



“The Real Faces of While Australia”

What we ended up with was a new way of seeing and understanding the records — not as the remnants of bureaucratic processes, but as windows onto the lives of people. All the faces are linked to copies of the original certificates and back to the collection database of the National Archives. So this is also a finding aid. A finding aid that brings the people to the front.

According to Margaret Hedstrom, the archival interface ‘is a site where power is negotiated and exercised.’^[1] Whether in a reading room or online, finding aids or collection databases are ‘neither neutral nor transparent’, but the product of ‘conscious design decisions’. We would like to think that this interface gives some power back to the people within the records. Their photographs challenge us to do something, to

think something, to feel something. We cannot escape their discomfiting gaze.

But this interface represents another subtle shift in power. We could create it without any explicit assistance or involvement by the National Archives itself. Simply by putting part of the collection online, they provided us with the opportunity to develop a resource that both extends and critiques the existing collection database. Interfaces to cultural heritage collections are no longer controlled solely by cultural heritage institutions.

It's these two aspects of the power of interfaces that I want to focus on.

There is a growing number of examples where the records created by repressive or discriminatory regimes have, in Eric Ketelaar's words, 'become instruments of empowerment and liberation, salvation and freedom.'^[2] Nazi records of assets confiscated during the Holocaust have been used to inform processes of restitution and reparation. Government records have helped members of Australia's Stolen Generations trace family members. Descendants of inmates incarcerated by American colonial authorities in what was the world's largest leprosy colony in the Philippines, have embraced the administrative record as an affirmation of their own heritage and survival.^[3] Records can find new meanings. Power can be reclaimed.

Technology can help. [Tim Hitchcock](#) has described how something as simple as keyword searching can turn archives on their heads. Recordkeeping systems tend to reflect the structures and power relations of the organisations that create them. The 'hierarchical and institutional nature of most archives', Hitchcock argues, 'contains an ideological component which is sucked in with every dust-filled breath.'^[4] But digitisation and keyword searching free us from having to follow the well-worn paths of institutional power. We can find

people and follow their lives against the flow of bureaucratic convenience. We can gain a wholly new perspective on the workings of society. 'What changes', Hitchcock asks, 'when we examine the world through the collected fragments of knowledge that we can recover about a single person, reorganised as a biographical narrative, rather than as part of an archival system?'^[5]



"Unknown No Longer"

Projects such as [Unknown no longer](#) may help us answer that question.

It's aiming to extract the names and biographical details of slaves from the 8 million manuscript documents held by the Virginia Historical Society. The documents include court records, receipts, wills and inventories. Here is a page from the 'Inventory of Negroes at Berry Plain Plantation, King George County, Virginia' for 1855, listing names, occupations and valuations.

Tim Hitchcock is one of the directors of [London Lives](#) a project that similarly seeks to find the people in 240,000 manuscript pages documenting the lives of plebeian Londoners in the 17th century.

**Inventory of Negroes
at Berry Plain
Plantation, King
George County,
Virginia**

<http://unknownlonger.vahistorical.org/record/92/38>



More than three million names have already been extracted from the records of courts, workhouses, hospitals and other institutions. Work is continuing to link these names together, to merge these various shards of identity and piece together the experiences of London's poorest inhabitants.



<http://www.londonlives.org/>

“London Lives”

[Remember me](#), from the US Holocaust Memorial Museum, is working with photographs taken by relief agencies in the aftermath of World War Two. The photographs are of displaced children who survived the

Holocaust but were separated from families. What happened to them? The project is seeking public help to identify and trace the children.



<http://rememberme.ushmm.org/>

“Remember Me?” – The United States Holocaust Memorial Museum

These are all projects about finding people. They are projects about finding the oppressed, the vulnerable, the displaced, the marginalized and the poor and giving them their place in history. This is what Kate and I hope to do with [Invisible Australians](#), the broader project of which our faces experiment is part.



<http://invisibleaustralians.org/>

“Invisible Australians”

'Invisible Australians' aims to extract more than just photographs. We want to record and aggregate the biographical data contained within the records of the White Australia Policy to extract the data and rebuild identities.



National Archives of Australia: ST84/1, 1905/61-70

But [we want to do more](#); we want to link these identities up with other records, with the research of family and local historians, with cemetery registers and family trees, with newspaper articles and databases we don't even know about yet. We want to find people, families, and communities.

It's ridiculously ambitious and totally unfunded. But it is possible.

The most exciting part of online technology is the power it gives to people to pursue their passions. As with the faces, we don't need the help of the National Archives. We need the records to be digitized, but that's happening anyway and we can afford to be patient. Most of the tools we need already exist, and are free. In the past twelve months, for example, there have been a number of open source tools released for crowd-sourced transcription of manuscript records.

People with passions, people with dreams, people who are just annoyed and impatient, don't have to wait for cultural institutions to create exactly what they need. They can take what's on offer and change it.



Before and After Interface Modification



One of the display possibilities with the custom interface

Interfaces can be modified. It is amazingly easy to write a script that will change the way a web page looks and behaves in your browser. I was frustrated by the standard interface to digitized files in the National Archives of Australia's Recordsearch database—so I changed it.

Not only did I make it look a bit nicer, I added new functions. My script lets you print a whole file or a range of pages and display the entire contents of the file on a pretty cool three-dimensional wall.

I've shared this script, and [a few other Recordsearch enhancements](#). Anyone can install them with a click and use them.



[WraggeLabs Emporium](#)

Interfaces are sites of power, and we can claim some of that power for ourselves. Online technologies not only free us from the having to brave the physical intimidation of the reading room, they free us up to engage with the records in new ways. The archivist-on-duty would probably not be pleased if I pulled out some scissors and started snipping photos out of certificates. Or if I pulled a file apart and pasted its contents on the wall. But online we are free to experiment.

The power of cultural heritage organisations is perhaps expressed most forcefully in their ability to control the arrangement and description of their collections. 'Every representation, every model of description, is biased', note Verne Harris and Wendy Duff, 'because it reflects a particular world-view and is constructed to meet specific purposes.'^[6] Archives, libraries and museums are already starting to share this power, by allowing tagging, or seeking public assistance with description through crowd sourcing projects. But most of these activities still happen within spaces created and curated by the institutions themselves. Our cathedrals of culture might be opening their doors and inviting the public to participate in their ceremonies, but that doesn't make them bazaars. The architecture still speaks of authority.

In any case, people already have a space where they can explore and enrich collections — it's called the internet.

It would be great to see cultural institutions doing more to watch, understand and support what people are doing with collections in their own spaces — following them as they pursue their passions, rather than thinking of ways to motivate them.

A quick example... You might have heard of [Zotero](#), it's an open source project that lets you capture, annotate and organize your research materials.

One cool thing about Zotero is that you can build and contribute little screen scrapers, called translators, that let Zotero extract structured data from any old collection database. You might not be surprised to learn that I've created a translator for Recordsearch. Another cool thing about Zotero is that you can share the stuff that you collect in public groups.



[“Invisible Australians” Zotero Group](https://www.zotero.org/groups/invisible_australians)

Put those two cool things together and what do you have? Well to me they spell out user generated finding aids — parallel collection databases created by researchers simply pursuing their own passions.

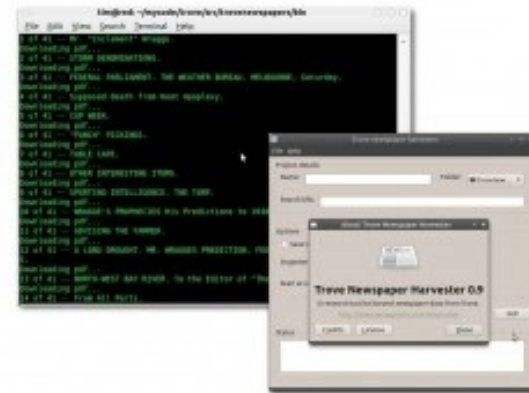
Linked Open Data greatly increases opportunities for collection description to leak into the wider web. If objects and documents are identified with a unique URL, then anyone can make and publish statements about them in machine-readable form. These statements can then be aggregated and explored. Initiatives such as the [Open Annotation Collaboration](#) will hasten the development of these shared descriptive and interpretative layers around our cultural collections.

And of course all this descriptive and interpretative work can be harvested back to enhance existing collection databases. We could start doing it now.

As well as exploring the possibilities of user-generated content, cultural institutions are starting to open up their collection data for reuse. APIs are great (though Linked Open Data is better), and New Zealand is lucky to have an organisation like [DigitalNZ](#) which just gets it. People can and will make cool things with your stuff.

But again, we don't have to wait for everything to be delivered in a convenient, machine-readable form. If it's on the web anybody can scrape, harvest and experiment.

You may know about the [National Library of Australia's newspaper digitisation project](#)—it's building a magnificent resource. But I wanted to do more than just find articles. I wanted to explore and analyze their content on a large scale. So I built a screen scraper to extract structured data from search results, and then used the scraper to power a series of tools. I have a [harvester](#) that lets you download an entire results set—hundreds or thousands of articles—with metadata neatly packaged for further analysis.



[“Query Harvester” by WraggeLabs Emporium](#)

Or what about a [script](#) that graphs the occurrence of search terms over time, and allows you to ask questions like [When did the Great War become the First World War?](#)

In the end I got a bit carried away and built my own [public API](#) to the Trove newspaper database.

the collections themselves squarely at the centre of our thoughts and actions. Instead of concentrating on the relationship between the institution and the public, we can focus on the relationship we both have with the collections.

It's all about the stuff.

It's all about the respect and responsibility we both have for our collections.

It's all about the respect and responsibility we both have for people like this.

Originally published by Tim Sherratt on [December 1, 2011](#). Revised for the *Journal of Digital Humanities* March 2012.

This is a modified version of a [paper I presented](#) at the [National Digital Forum](#), 30 November 2011.

Notes:

[1] Margaret Hedstrom, "Archives, Memory, and Interfaces with the Past," *Archival Science* 2 (2002): 22.

[2] Eric Ketelaar, "Archival Temples, Archival Prisons: Modes of Power and Protection," *Archival Science* 2 (2002): 229.

[3] Ricardo L Punzalan, "Archives and the Segregation of People with Leprosy in the Philippines," presented at *ICHORA-4*, University of Western Australia, 2008, pp. 93-102.

[4] Tim Hitchcock, "Digital Searching and the Re-formulation of Historical Knowledge," in Mark Greengrass and Lorna Hughes (eds), *The Virtual Representation of the Past* (Farnham, UK: Ashgate, 2008), 87.

[5] Tim Hitchcock, p.90.

[6] Wendy M. Duff and Verne Harris, "Stories and Names: Archival Description as Narrating Records and Constructing Meanings," *Archival Science* 2 (2002): 275.

MOYA Z. BAILEY

All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave

Following a [fascinating talk](#) by Ed Finn on the changing role and source of literary criticism in a digital age, Natalia Cecire queried the implicit neutrality of a term like “nerd.” Melissa Harris-Perry’s [reclamation](#) aside, the racialized and gendered aspects of nerddom, and by extension the digital humanities, offer opportunities for a more explicit engagement with positionalities that lead “[white men to feel embattled](#).” How do those outside the categories white and male navigate this burgeoning disciplinary terrain?

The ways in which identities inform both theory and practice in digital humanities have been largely overlooked. Those already marginalized in society and the academy can also find themselves in the liminal spaces of this field. By centering the lives of women, people of color, and disabled folks, the types of possible conversations in digital humanities shift. The move “[from margin to center](#)” offers the opportunity to engage new sets of theoretical questions that expose implicit assumptions about what and who counts in digital humanities as well as exposes structural limitations that are the inevitable result of an unexamined identity politics of whiteness, masculinity, and ablebodiedness.

What counts as a digital humanities project? As an undergrad, I interacted with people who were actively doing intersectional digital humanities work in all but name in other arenas of the academy. Dr. Carla Stokes wrote her dissertation on the online culture of Black girls. She discussed how Black girls were using digital platforms like chat rooms, web pages, and blogs to create identity. Through the creation of the non-profit [Helping Our Teen Girls](#), Stokes offered an alternative online network (which she built) that was peer moderated to help address issues of cyber-bullying, and the targeting of youth online by adults. Stokes work is lauded in Girls Studies and Critical Media Studies. While certainly a digital humanities project, her work has not been legible as such.

In attempting to speak to and reach communities we felt accountable to outside academia, the [Crunk Feminist Collective](#) began blogging in 2010. A collective of about ten academics and activists use social media platforms to talk about the realities of our world in accessible feminist language. This hybridizing of cultural production and a theoretical praxis, falls outside the purview of mainstream digital humanities but has been utilized in classrooms across the country.

Scholars like [Lisa Nakamura](#) brilliantly bridge both cultural criticism and digital humanities. In her recent scholarship, Nakamura examines the exploitation of indigenous women’s labor in the construction of digital devices. Far from saying people of color are not engaged in digital humanities, Nakamura’s work begs for a recentering of the conversation on the parts of the field that are [messy](#). There is a need to address the complexities of globalization, colonization, and the alienated labor of people of color in the production of technology that advances digital scholarship practices that they will not be able to access or directly benefit from.

How and where do the humanities enter into digital humanities, and how can they change the way we talk in the field? I've been reading a lot of digital humanities blogs and the use of ableist language show that the work of disability scholars, [while tangentially acknowledged](#), may not have shifted practices. Words, like "lame," "stupid," and "retarded" are used to describe problematic elements in the field without any recognition of their own problems. In doing the work of creating and utilizing digital tools for better digital humanities projects, shouldn't we also be engaging the humanities themselves? There has been much needed criticism leveled at the free programming interface [Codecademy](#). Had the creators worked with scholars in educational studies might they have produced a more accessible learning tool? In re-imagining what counts as digital humanities, we can draw on the wisdom of scholars who have addressed related issues in their own fields of study. Of all the emergent interdisciplinary spaces, digital humanities is uniquely poised to apply academic research to itself and its products.

In blog posts, [Miriam Posner](#) and [Bethany Nowviskie](#) have both addressed the structures that impede women from connecting to digital humanities. The increase of women in higher level positions within universities have led to changes in the infrastructure, with child care and nursing nests cropping up on campuses across the country. Similarly, people of color have been engaging in [critical university studies](#) long before the 1990s when the field is said to have emerged. By demanding space as students and faculty, in addition to advocating for rights as the laborers that built and maintain these institutions, people of color have organized through concerted effort to bring about changes in institutional culture and structure.

As more diverse groups of people have entered the academy and the field of digital humanities, the contours have been redefined. We are sometimes the square pegs that expose the unacknowledged round

holes. There is an elasticity to digital humanities that makes this a solvable problem, and people are already working through it. The activism of groups like [#transformDH](#), the promise of [THATCamp Theory](#), and the work of [Critical Code Studies](#) are challenging the hacking through more directed yacking. Sparked by the dearth of women in the field, a [THATCamp Feminisms](#) has been proposed. Initiatives like [Black Girls Code](#) are truly grassroots, reaching girls of color in elementary and middle school with opportunities to engage STEM before they are [tracked](#) away from it.[1]

There is still a need to challenge the "add and stir" model of diversity, a practice of sprinkling in more women, people of color, disabled folks and assuming that is enough to change current paradigms. This identity based mixing does little to address the structural parameters that are set up when a homogeneous group has been at the center and don't automatically engender understanding across forms of difference. It elides the scholarship already in production that may not be readily apparent when looking from a singular perspective. As opposed to meeting people where they are, where people of color, women, people with disabilities are already engaged in digital projects, there's a making of room at an already established table. Work that is already aligned with the digital humanities and perhaps even pushing the field in new directions should be celebrated and sought out, a process that will no doubt reveal, that some of us are brave.

* Title adapted from [All the Women Are White, All the Blacks Are Men, But Some of Us are Brave](#). [2]

Notes:

[1] Karolyn Tyson, *Integration Interrupted: Tracking, Black Students, and Acting White after Brown* (New York: Oxford University Press, 2011).

[2] Gloria T. Hull, Barbara Smith, and Patricia Bell Scott, eds. *But Some of Us are Brave* (Old Westbury, N.Y.: The Feminist Press, 1983).

Reviews

Review of WordSeer, produced by Aditi Muralidharan, Marti Hearst, and Bryan Wagner

Amy Earhart 106

Review of Bookworm, produced by Benjamin Schmidt, Martin Camacho, et al

Boone B. Gorges 109

Review of QueryPic, produced by Tim Sherratt

Jeremy Boggs, David McClure, Eric Rochester, and Wayne Graham 112

AMY EARHART

Review of WordSeer,

produced by Aditi Muralidharan, Marti Hearst, and Bryan Wagner

The [WordSeer](#) tool, developed at the University of California, Berkeley by [Aditi Muralidharan](#) and [Marti Hearst](#) with research partner [Bryan Wagner](#), is an exploratory analysis or “sensemaking” environment for literary texts. The tool is based on an understanding of literary analysis as a cyclical, rather than a linear, process, a notion that has been underemphasized in tool development where visualizations and datamining have generally been seen as exposing the text for scholarly treatment. WordSeer allows you to read a text, search for relationships between words and phrases, examine grammatical relationships, and examine produced heat map and tree visualizations.

WordSeer is the only tool specifically designed for literary analysis that will perform grammatical searches using natural language processing, a crucial step forward in literary tool approaches. The code is open and the developers encourage others to reuse and modify as needed. The tool accepts XML texts only, a reasonable choice given the prevalence of TEI/XML texts in digital literary scholarship though it would be nice to have the option of utilizing .txt files as well.

WordSeer is still in its infancy, so some issues should be resolved as it develops. Documentation has not yet been written, leaving the user to relying on experimentation for use. Several functions, including example searches and date selections, do not work consistently. In addition, the current version of WordSeer runs on only three sets of texts: The Slave Narratives from [Documenting the American South](#), Shakespeare, and Stephen Crane. The [Federal Writers Project Slave Interviews](#) is listed as a test set, but not yet available. The developers have plans to open the tool for general use in 2013.

While the tool is limited at this stage of production it shows great promise. Grammatical searches using natural language processing promise a greater flexibility for scholars interested in moving from the macro to micro level of analysis. One of the most useful features of the tool is the ability to modify results of searches from the word or phrase level, allowing the scholar to start with distance reading and, based on results, drill down in to the materials for greater analysis.

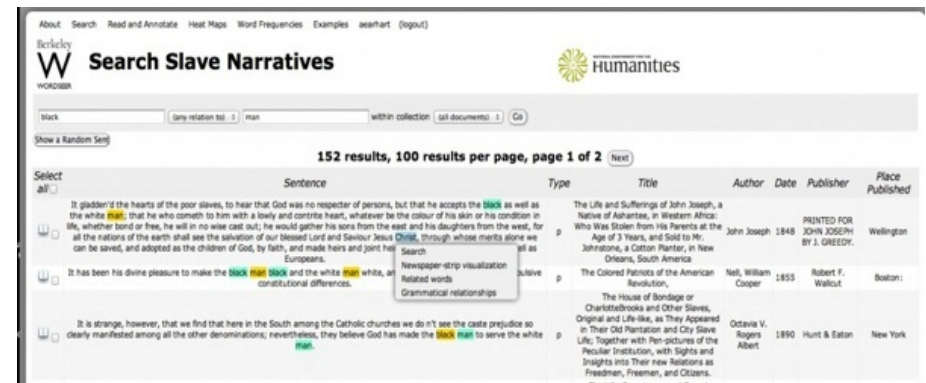


Image 1: Modification of Search Results From Within Text

In this respect the tool fulfills its claim to create a cyclical environment for scholarly exploration. One concern with grammatical structure

Given the textual diversity, it is impossible to prove or disprove Olney's criteria, which is premised on the black-authored autobiography. The same misunderstanding of the dataset is apparent with the decision to split selection for time periods at 1838. 1838 is the year slavery was abolished in the UK, but this set of data is focused on North America. I would encourage the WordSeer team to develop stronger ties to literary scholars who are able to move between technology and literary scholarship to develop more robust literary questions for analysis.

Regardless of such concerns, the tool offers great promise, and I await the open release in 2013.

Review of Bookworm, produced by Benjamin Schmidt, Martin Camacho, et al

[Bookworm](#) is a tool that allows users to create visualizations charting the use of words or phrases in selected large corpora over specified periods of time. The software was developed by a group of researchers at the [Harvard University Cultural Observatory](#) [ref]Ben Schmidt, Martin Camacho, Neva Cherniavsky, Erez Lieberman Aiden, and Jean-Baptiste Michel.[/ref] as a follow-up to the 2010 project that resulted in [a cover story in Science](#), the [Google Books Ngram Viewer](#), and the coining of the term '[culturomics](#)'. Bookworm development is still in its alpha phase, but already the software shows great promise as a tool for scholarly exploration of historical trends in large collections of books.

The purpose of Bookworm is to track the frequency with which a phrase is used over a certain time span, within user-defined subsets of Bookworm's total book collection. Users can carve out subcollections in a number of different ways: by geography (where the book was published), language, Library of Congress classification, and date (publication date, date of authorship, birthdate of author). These criteria can be combined, and compared against each other, to nice effect. For example, we can compare the frequency of the name 'Simon Cameron' as it appears in all books, versus all books published in

Pennsylvania, versus Pennsylvanian books under the subject heading "History of the Americas" [Figure 1].

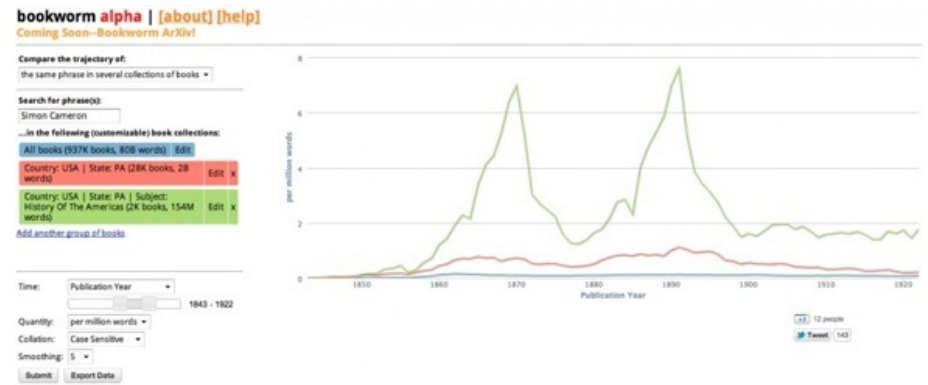


Figure 1: Frequency of “Simon Cameron” in books with “History of the Americas” subject Heading

The results are, in this case, predictable: Cameron's name appears more frequently in publications from his home state, with peaks in the decade or so after the Civil War. But this rather mundane example gestures toward the potential for richer exploratory searches across custom-defined subcollections of Bookworm's index.

Similarly, we can compare the frequency of different phrases across a single corpus. For instance, the phrase 'Simon Cameron' is (again, unsurprisingly) more frequent in Pennsylvania books than 'Gideon Welles' [Figure 2].

The power of a tool designed for the exploration of a specific corpus is, in large part, dependent on the size and quality of that corpus. Bookworm's index is based on the public domain works available through [Open Library](#), which numbers nearly one million books. The public domain limitation is notable for a few reasons. For one thing, it means that Bookworm's corpus is more limited than that of the Google

Books project, both in terms of sheer size (the original *Science* paper cites a corpus of over five million books) and in terms of usable date ranges (most US public domain works come from before 1922, the horizon for most works under current US copyright law). At the same time, limiting the collection to public domain works means that Bookworm can link to the full text of query results [Figure 3].

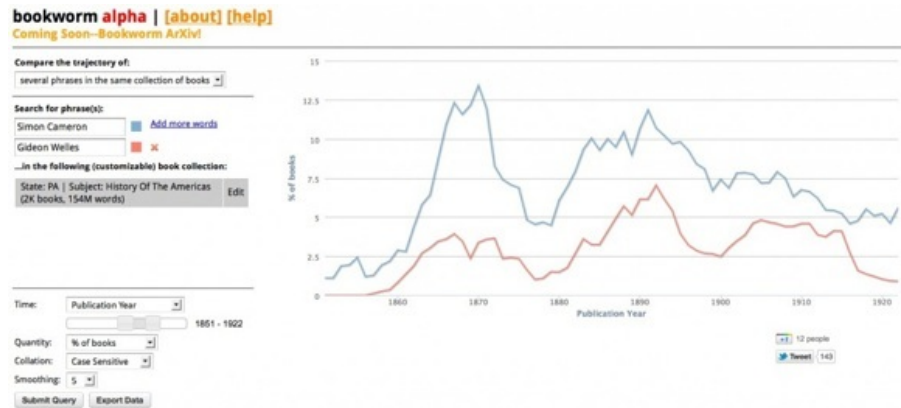


Figure 2: Frequency of “Simon Cameron” and “Gideon Welles” in books published in Pennsylvania with “History of the Americas” subject heading

In this way, Bookworm provides a link (both figurative and literal) between the very distant reading of macro-level quantitative analysis, and the close reading of specific texts that is crucial to contextualizing the qualitative results.

Setting aside limitations imposed by what copyright law *excludes* from the corpus, some of Bookworm's notable limitations come from the weaknesses of the texts that *are* in the collection. The OCR (optical character recognition) process used to scan the books is imperfect. As a simple example, the frequency of a non-word like 'hiftery' (for 'history') shows that the software has a hard time distinguishing characters like the medial 's'. And the metadata used by Open Library (and thus by

Bookworm) to categorize and filter collections is, in places, incomplete or incorrect. Scholars should take heart, however, that Open Library's metadata is publicly modifiable: users can add or edit a book's info, which Bookworm - and any other tool using Open Library data - will recognize the next time it refreshes its index.

Books for series Simon Cameron matching constraints in 1883

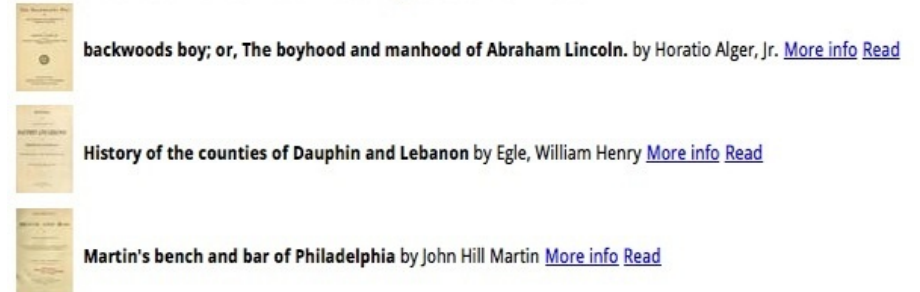


Figure 3: Books for series Simon Cameron matching constraints in 1883

The team behind Bookworm is at work on numerous improvements. Publicly available installations of Bookworm are being developed to track other large corpora; [Bookworm arXiv](#), just announced, draws from the scientific papers of [arXiv.org](#). The development team is [particularly interested](#) in conceptualizing Bookworm as an interface for browsing library catalogs. And perhaps most exciting is the prospect of an eventual general release of the software - including the visualization interface as well as the server-side tools necessary for indexing arbitrary collections of text - under a free software license. [ref] This possibility was conveyed to me by developer Ben Schmidt in an email exchange.[/ref] Such a release would allow individual scholars or other organizations to host their own Bookworm instances, connected to whatever arcana they see fit. This would be a most

welcome addition to the existing library of tools for macro text analysis.

Even in this early incarnation, Bookworm is an easy-to-use and powerful way for interested parties to get started with quantitative analysis of a large and important corpus of works in the public domain.

JEREMY BOGGS, DAVID MCCLURE, ERIC ROCHESTER, AND WAYNE GRAHAM

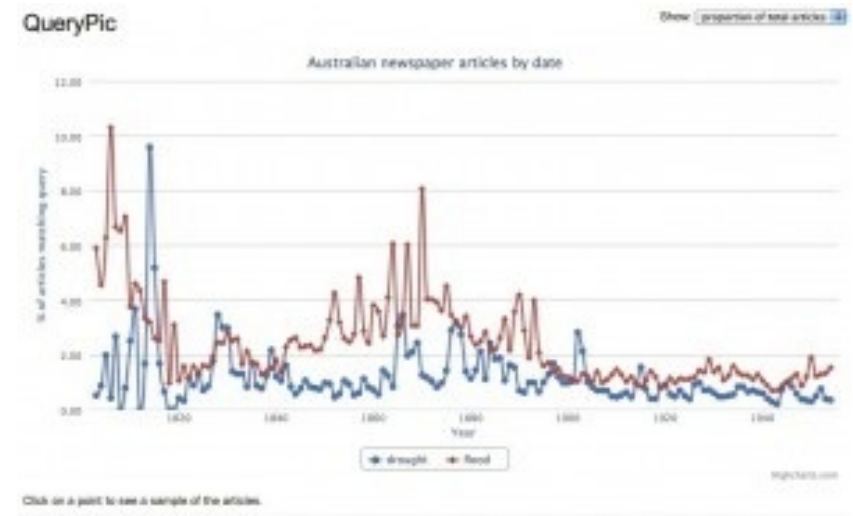
Review of QueryPic, produced by Tim Sherratt

[QueryPic](#) is a graphical search summarizer that mines content in the [Trove](#) newspaper archive from the [National Library of Australia](#). The program is part of Tim Sherratt's larger [TroveNewspaper software project](#), which allows researchers to obtain parsable data from the Trove collection. QueryPic searches the Trove newspaper archive, scrapes the returns, and graphs the results of the search. QueryPic uses the Python programming language, and is [available on GitHub](#) under a GPL v3 license.

Using QueryPic requires some comfort entering commands on the console or terminal. To start, users first must download the TroveNewspaper package (or clone the Git repository) to their computer. Then, users must open a terminal and change into the TroveNewspaper package directory, and type commands in order to submit a query to the Trove archive:

```
python do_totals.py "http://trove.nla.gov.au/newspaper/result?q=drought" -g "drought_flood"
```

For the example above, QueryPic generates a chart called "drought_flood" in a 'graphs' directory that plots the number of articles found by year for a query on "drought." The chart provides a great way to explore the results of any given query on the Trove newspaper archive. Clicking on each point on the graph reveals a list of articles returned by the query, with links to view each article in the Trove archive. Sherratt includes several options for generating the



A screenshot of a chart generated with QueryPic

graphs. Users can generate a single graph for multiple queries to compare the results of each, by making sure the value for the graph name '-g' is the same. Users also can plot changes in the query results by month instead of year.

The graphs are generated on a well-formatted HTML file, using [HTML5 Boilerplate](#) for page markup and the [960 Grid System](#) for styling and layout. QueryPic also uses the [HighCharts](#) JavaScript library to create the line graph. Transferring

these graphs to a public web page would be relatively easy, as long as the accompanying CSS and JavaScript files are also transferred.

A few areas for improvement for the scripts themselves: The way the TroveNewspaper package is currently written sends a high volume of requests to the Trove archive, so the Australian National Archives could potentially throttle requests if increased use of the script occurs. Perhaps the TroveNewspaper package should limit the rate at which it issues requests to avoid overtaxing the National Archives system. The project might also benefit from having some continuous testing that will parse the data from the Trove archive, which could provide regular alerts if the Trove site changes its markup and consequently breaks any scripts in the TroveNewspaper package. Also, the way that the software is currently constructed makes it difficult for developers to use this in a larger application. However, adding a setup file to the package would allow the software to be easily used in other Python applications.

Of course, any script that relies on screen scraping will eventually break. Even slight changes to the HTML you are scraping can break the script, as was the case for the review team. Fortunately the fix was easy, and the team member who discovered the issue (Rochester) [submitted a fix](#), which Sherratt quickly added to the TroveNewspaper package. This quick response, combined with Sherratt's own commit history to the project and [exploration of its use](#) on his blog, indicates that development on the project is active and attentive.

QueryPic, and the TroveNewspaper package of which it is part, are a great example of how digital humanists are exploring and implementing new functionality for existing resources. QueryPic is of particular use to anyone using the Trove archive to research Australian history and culture. Sherratt's broader software projects also serve as a model for anyone with moderate programming knowledge (or access to

someone who does) who wants to create tools for use with other archives.

Contributors

Moya Bailey is a graduate candidate at Emory University where she explores critical race, feminist, and disability studies. Her current work focuses on constructs of health and normativity within a US context. She is interested in how race, gender, and sexuality are represented in media and medicine. She is a blogger and digital alchemist for the Crunk Feminist Collective. In a co-authored piece for *Ms. Magazine*, Bailey proclaims "Black Feminism Lives (online)!" and chronicles the digital discourses of race, gender, and politics as articulated by young black women in cyber space.

Jean Bauer is the Digital Humanities Librarian at Brown University, where she works with faculty, students, and fellow librarians to design and implement digital systems that showcase and facilitate scholarship in the humanities. Through a combination of formal training and curiosity she is an early American historian, database designer, and

photographer. She is finishing her dissertation, "Revolution-Mongers: Launching the U.S. Foreign Service, 1775-1825," in the Corcoran Department of History at the University of Virginia and has developed [The Early American Foreign Service Database](#). For more information, see her website www.jeanbauer.com.

Chad Black is an Assistant Professor of Latin American History at the University of Tennessee, Knoxville, and author of *The Limits of Gender Domination: Women, the Law, and Political Crisis in Quito, 1765-1830*. In response to the abundance offered by digitizing approaches to the Colonial Latin American Archive, my research has increasingly turned to humanities computing.

Jeremy Boggs is Design Architect for the the University of Virginia's Scholars' Lab. He is currently ABD in History at George Mason University, where he is writing a dissertation on design methodologies for digital history.

Alison Booth is Professor of English at the University of Virginia, where she teaches narrative and Victorian literature. She is the author of *Greatness Engendered: George Eliot and Virginia Woolf* (Cornell, 1992) and *How to Make It as a Woman: Collective Biographical History from Victoria to the Present* (Chicago, 2004) and editor of *Wuthering Heights* (Longman) and *Introduction to Literature* (Norton). Her research explores both the audience reception of authors and biography of or in groups, in an Anglo-American context. She is currently engaged in a digital project, [Collective Biographies of Women](#), and a book on literary tourism, biography, and house museums, "Writers Revisited."

Dr. Peter Bradley is Associate Professor of Philosophy at McDaniel College in Westminster, MD. Prior to coming to the Digital Humanities, he worked primarily on the Philosophy of Perception, specifically the Philosophy of Psychophysics. In the past few years, most of his thinking has focused on the Philosophy of Liberal Education. In his spare time, he develops technological resources for teaching Logic and Critical Thinking. He's published articles in *Teaching Philosophy*, *Philosophical Psychology*, and the *Journal of Philosophy*.

Natalia Cecire is a postdoctoral fellow at the Fox Center for Humanistic Inquiry at Emory University, and a managing editor at [Arcade](#).

Dan Cohen is an Associate Professor in the Department of History and Art History at George Mason University and the Director of the Roy Rosenzweig Center for History and New Media. His own research is in European and American intellectual history, the history of science (particularly mathematics), and the intersection of history and computing. He is the co-author of *Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web* (University of Pennsylvania Press, 2005), author of *Equations from God: Pure Mathematics and Victorian Faith* (Johns Hopkins University Press, 2007), and has published articles and book chapters on the history of mathematics and religion, the teaching of history, and the future of history in a digital age in journals such as the *Journal of American History*, the *Chronicle of Higher Education*, and *Rethinking History*. Dan is an inaugural recipient of the American Council of Learned Societies' Digital Innovation Fellowship.

Marc Downie and **Paul Kaiser** are two of the three artists in OpenEndedGroup, which has pioneered a new approach to digital art that frequently combines three signature elements: non-photorealistic 3D rendering; the incorporation of body movement by motion-capture and other means; and the autonomy of artworks directed or assisted by artificial intelligence. OpenEndedGroup artworks span a wide range of forms and disciplines, including dance, music, installation, film, and public art. They create their work by means of their own extensive software platform, Field. Released as open source, Field has received support from the Mellon Foundation, the NEH, and the NSF. Downie, whose background is in music and computer science, has a PhD from the Media Lab MIT and an MA from Cambridge University. Kaiser's background is in film, writing, and special education; he has a BA from Wesleyan University and an MEd from American University. Among the prizes they have won individually or collectively are a Guggenheim Fellowship, the John Cage Award from the Foundation for Contemporary Arts, a Media Arts Fellowship from the Rockefeller Foundation, and a Bessie award for the BIPED décor. Their venues have included Lincoln Center, the New York Film Festival, the Barbican Center, the Hayward Gallery, ICA Boston, Sadler's Wells, the Festival d'Automne, the Sundance Film Festival, and many others.

Amy Earhart is an Assistant Professor in the Department of English at Texas A&M University. Her work has appeared in *DHQ: Digital Humanities Quarterly*, the *Chronicle of Higher Education/ProfHacker*, and *Debates in Digital Humanities* (Minnesota 2010), among other venues. She has co-edited a collection of essays titled *The American Literature Scholar in the Digital Age* with Andrew Jewell (Michigan 2010). Earhart is at work on a monograph titled "Traces of the Old, Uses of the New: The Emergence of the Digital Humanities." Her digital projects include the development of the 19th-Century

Concord Digital Archive in partnership with the Concord Free Public Library.

Michael Edson is the Smithsonian Institution's Director of Web and New Media Strategy. Michael has worked on numerous award-winning projects and has been involved in practically every aspect of technology and New Media for museums, including content development, digitization, blogging, gaming, public access to collections, information architecture, networking, place-of-business applications, programming, project management, graphic design, animation, audio and video production, mobile platforms, and citizen-created content. In addition to developing the Smithsonian's first [Web and New Media Strategy](#), the Smithsonian Commons concept, and the Smithsonian's multi-award winning [Web and New Media Strategy Wiki](#), Michael helped create the Smithsonian's first blog, [Eye Level](#), and the first Alternative Reality Game to take place in a museum, [Ghosts of a Chance](#). Michael is an O'Reilly Foo Camp veteran and was named a Tech Titan 2011: person to watch by *Washingtonian* magazine. He has a BA from Wesleyan University.

Fred Gibbs is an Assistant Professor in the Department of History and Art History at George Mason University and Director of Digital Scholarship at the Roy Rosenzweig Center for History and New Media. His current research explores late medieval and early modern understandings of poison, as well as the role of critical theory in the digital humanities. He teaches courses on digital history and various topics in the history of science, medicine, and technology.

Boone B. Gorges is a freelance web developer and software consultant based in Brooklyn, NY. He is the Lead Developer and Director of Research Projects for the [CUNY Academic](#)

[Commons](#) and [Commons In A Box](#). He specializes in the development of WordPress plugins, and is on the core development team of such WordPress-related project as [BuddyPress](#) and [Anthologize](#). Gorges holds an MPhil in Philosophy from the CUNY Graduate Center.

Wayne Graham is Head of the R&D team at the Scholars' Lab. He studied US Colonial History at [William and Mary](#), and wrote a couple of books on [developing applications for the Facebook platform](#). His research interests include computer graphics, augmented reality, architectural history, quantitative and digital methodology, as well as frontier studies.

Mitch Green is the NEH/Horace Goldsmith Distinguished Teaching Professor of Philosophy in the Department of Philosophy at the University of Virginia. He has held fellowships from the National Endowment for the Humanities, the National Science Foundation, the Andrew Mellon Foundation, the Virginia Foundation for the Humanities, the National Humanities Center, the Mead Endowment, the University of Pittsburgh's Center for Philosophy of Science, the Center for Contemplative Mind in Society, the American Council of Learned Societies, the University of Virginia's Teaching Resource Center, and the University of Virginia's Shannon Center for Advanced Studies. His specializations are in Philosophy of Language, Philosophy of Mind, and Aesthetics. He is also interested in Metaphysics, Decision Theory, the Theory of Action, and the history of analytic philosophy. His current research interests include evolutionary biology of communication, speech acts and their role in communication, empathy, self-knowledge, self-expression, and attitude ascription.

Tim Hitchcock is Professor of Eighteenth-Century History at the University of Hertfordshire. In collaboration with Robert Shoemaker and others he has created a series of digital resources providing scholarly and public access to the sources that underpin the history of eighteenth-century London, including the [Old Bailey Online](#), [London Lives](#), [Connected Histories](#), and most recently [Locating London's Past](#). Together, these sites make available some 12 billion words of digitized content. With a range of collaborators including Bill Turkel at the University of Western Ontario, he has also been at the forefront of the process of applying the tools of digital humanities, of text and data mining, to the analysis of these new resources. The author of ten books on the histories of eighteenth-century poverty, sexuality and masculinity, Hitchcock has degrees from the University of California at Berkeley, and Oxford University. He blogs at [Historyonics](#), and as @TimHitchcock on Twitter.

Nik Honeysett is Head of Administration for the J. Paul Getty Museum in Los Angeles, administrating the Getty's two campuses in Brentwood and Malibu. Formerly, he managed the Getty's Web Group and before moving to Los Angeles, he was Head of Production at Cogapp, a UK-based consultancy specializing in interactive and online multimedia for the cultural sector in Europe and the USA. He is a former chair of the American Association of Museum's Media and Technology Standing Professional Committee and currently sits on AAM's Board of Directors. His hobbies include writing short summary paragraphs about his career and referring to himself in the third person. Blog: www.musematic.net / midea.nmc.org Twitter: @nhoneysett

Anne Houston is the Director of Humanities and Social Sciences at the University of Virginia Library.

Alexis Lothian recently completed her PhD in English and Gender Studies at the University of Southern California. Her book manuscript, "Deviant Futures: Queer Temporality and the Cultural Politics of Science Fiction," reorients recent scholarship on queerness and futurity by approaching it through histories of speculative fiction and media by queers, feminists, and people of color in twentieth-century Britain and America. She also works on digital media, particularly contemporary transmedia cultural production and online culture, and is developing a second book project on the intersections between online science fiction fan communities and social justice politics. Her work has been published in *Cinema Journal*, *Camera Obscura*, *Extrapolation*, *English Language Notes*, *Transformative Works and Cultures*, and *Social Text Periscope*, among other venues. As a core member of the #transformDH collective, she is developing several projects aimed at bringing together queer and ethnic studies critique with the methodologies of the digital humanities. She has been a member of the editorial team for the open access journal *Transformative Works and Cultures* since its founding in 2008. Her website is queergeektheory.org.

David McClure is Web Applications Specialist on the Scholars' Lab R&D team. He graduated from Yale University with a degree in the [Humanities](#) in 2009, and prior to joining the SLab, worked as an independent web developer and communications consultant in San Francisco, New York, and Madison, Wisconsin.

Elijah Meeks is the Digital Humanities Specialist at Stanford University, where he helps to bring network analysis, text analysis, and spatial analysis to bear on traditional humanities research questions. He has worked as the technical lead on The Digital Gazetteer of the Song Dynasty, Authorial London, and ORBIS: The Stanford Geospatial

Network Model of the Roman World. In his time at Stanford, he's worked with Mapping the Republic of Letters, the Stanford Literary Lab, and the Spatial History Lab, as well as individual faculty and graduate students, to explore a wide variety of digital humanities research questions.

Patrick Murray-John holds a PhD in medieval literature from the University of Wisconsin - Madison. After contingent faculty roles at the University of Mary Washington and the University of Richmond, he became an Instructional Technology Specialist at the University of Mary Washington. There, he began his journey into digital humanities and coding. He now works on various projects, mostly Omeka-related, and manages the Omeka development team at the Roy Rosenzweig Center for History and New Media.

Trevor Owens is a Digital Archivist with the National Digital Information Infrastructure and Preservation Program (NDIIPP) in the Office of Strategic Initiatives at the Library of Congress. He is finishing a PhD in social science research methods and instructional technology in the Graduate School of Education at George Mason University. Trevor wrote a book for Arcadia Press about the history of Fairfax County told through postcards, and published articles in the journals *Simulation & Gaming*, *Science Communication*, *Cultural Studies of Science Education*, and *On the Horizon*. In 2009 he was chosen by Read Write Web as one of the 50+ Semantic Web Pros to Follow on Twitter, and won the CW Bright Pixel Prize for building *PlayingHistory*, an open collaborative directory of digital historical games and interactives.

Stephen Ramsay is an Associate Professor in the Department of English and Fellow in the Center for Digital Research in the

Humanities at the University of Nebraska – Lincoln. He splits his time between pontificating about Digital Humanities, teaching humanities majors to program, undertaking analysis and visualization of text corpora, and designing and building text technologies for humanist scholars. He is the author of *Reading Machines: Toward an Algorithmic Criticism* (University of Illinois Press, 2011).

Eric Rochester is Senior Developer on the Scholars' Lab R&D team. He has a PhD from the University of Georgia in English. Before joining the Scholars' Lab, he worked as a consultant, programmer, and computational linguist for a number of technology firms, and consulted with the Oxford University Press and the University of Georgia's Linguistic Atlas Projects.

Mark Sample teaches and researches contemporary literature and new media in the Department of English at George Mason University. A vocal practitioner and critic of the digital humanities, Professor Sample has contributed to *Hacking the Academy* (University of Michigan Press) and *Debates in the Digital Humanities* (University of Minnesota Press). He also has a collaboratively written book forthcoming from MIT Press, which combines critical code studies, software studies, and platform studies, as well as more traditional forms of literary and textual scholarship to understand creative computing on the Commodore 64. Professor Sample is a regular contributor to ProfHacker, a feature at the *Chronicle for Higher Education* that focuses on pedagogy and scholarly productivity. He also writes for Play the Past, a collaboratively edited scholarly blog that explores the intersection of cultural heritage and games.

Tom Scheinfeldt is Managing Director of the Center for History and New Media and Research Assistant Professor of History in the

Department of History and Art History at George Mason University. Tom has lectured and written extensively on the history of popular science, the history of museums, history and new media, and the changing role of history in society. He has worked on traditional exhibitions and digital projects at the Colorado Historical Society, the Museum of the History of Science in Oxford, The Louisiana State Museum, the National Museum of American History, and the Library of Congress. In addition to managing general operations at the Center for History and New Media, Tom directs several of its online history projects, including Omeka, THATCamp, One Week | One Tool, the September 11 Digital Archive, the Hurricane Digital Memory Bank, the Papers of the War Department, 1784-1800, and Gulag: Many Days, Many Lives.

Benjamin M. Schmidt is the Visiting Graduate Fellow at the Cultural Observatory at Harvard University, and a PhD Candidate in history at Princeton University. He writes about digital humanities at [Sapping Attention](#).

Ryan Shaw is an assistant professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill. He recently received his Ph.D. from the University of California, Berkeley School of Information, where he wrote a dissertation on how events and periods function as concepts for organizing historical knowledge. He is also the author of the LODE (Linking Open Descriptions of Events) ontology, recently adopted by the UK Archives Hub for their Linked Data effort. He has been involved in a number of digital humanities projects through his work with the Electronic Cultural Atlas Initiative, including Bringing Lives to Light: Biography in Context (2006-2009), Context and Relationships: Ireland and Irish Studies (2007-2010), and Editorial Practices and the Web (2010-

ongoing). In a previous life, he worked as a software developer in Tokyo, Japan.

Tim Sherratt is a digital historian, web developer and cultural data hacker who's been developing online resources relating to archives, museums and history since 1993. He has written on weather, progress and the atomic age, and developed resources including Bright Sparcs, Mapping our Anzacs and The History Wall. He is currently working as a freelance troublemaker, as well as being a Harold White Fellow at the National Library of Australia, a Director's Fellow at the National Museum of Australia, and an Adjunct Associate Professor in the Digital Design and Media Arts Research Cluster at the University of Canberra. Tim is one of the organisers of THATCamp Canberra and a member of the interim committee of the Australasian Association for the Digital Humanities.

Lisa Spiro works as director of [NITLE \(National Institute for Technology in Liberal Education\) Labs](#), with the liberal arts community to explore emerging technologies and develop collaborative approaches to integrating learning, scholarship and technology. Lisa has presented and published widely on the digital humanities, including contributions to [Debates in the Digital Humanities](#), [Collaborative Approaches to the Digital in English Studies](#), [#alt-academy: Alternate Academic Careers for Humanities Scholars](#), [The American Literature Scholar in the Digital Age](#), and *Digital Humanities Pedagogy* (under review). She is the founding editor of [Digital Research Tools \(DiRT\)](#) and authors the [Digital Scholarship in the Humanities](#) blog. Before coming to NITLE, Lisa directed the [Digital Media Center](#) at Rice University's Fondren Library and worked at the University of Virginia's Electronic Text Center. Lisa serves on the Executive Council for the [Association of Computers and](#)

[the Humanities](#), the Board of [DHCommons](#), and the Program Committee for the [Joint Conference on Digital Libraries](#).

William G. Thomas is the Angle Chair in the Humanities and Professor of History at the University of Nebraska-Lincoln. He is the author of *The Iron Way: Railroads, the Civil War, and the Making of Modern America* (Yale University Press, 2011). He held a Digital Innovation Fellowship in 2008 from the American Council of Learned Societies, and a National Endowment for the Humanities Digital Humanities Digging into Data research grant for "Railroads and the Making of Modern America." Thomas also served as the Visiting Professor of North American Studies at the Eccles Centre for American Studies at the British Library in London, England in 2008. He was the Director of the Virginia Center for Digital History at the University of Virginia from 1998-2005, and a co-editor of *The Valley of the Shadow: Two Communities in the American Civil War*.

Scott B. Weingart is an NSF Graduate Research Fellow and PhD student at Indiana University, where he studies Information Science and History of Science. His research focuses on the intersection of historiographic and quantitative methodologies, particularly as they can be used to study scholarly communications in the past and present. He also writes a blog called the [scottbot irregular](#), aiming to make computational tools and big data analytics accessible to a wider, humanities-oriented audience. When not researching, Scott fights for open access and the reform of modern scholarly communication.