



# Journal of Digital Humanities

VOL. 1 NO. 3 SUMMER 2012

DANIEL J. COHEN, EDITOR

JOAN FRAGASZY TROYANO, EDITOR

SASHA HOFFMAN, ASSOCIATE EDITOR

JERI WIERINGA, ASSOCIATE EDITOR

ISSN 2165-6673

---

CC-BY 3.0

---



**A PRESSFORWARD  
PUBLICATION**

ROY ROSENZWEIG CENTER FOR HISTORY AND NEW MEDIA

GEORGE MASON UNIVERSITY

# The Difference the Digital Makes

So much of the content of digital humanities begins in the analog world: documents that are scanned and indexed; maps that are recast in GIS; quantities that are converted to machine-readable tables. Although we tend to focus on the final product — the digital construction viewed over the web — we remain cognizant of this transition that artifacts of human expression have taken.

In this issue of the *Journal of Digital Humanities*, several scholars take a deeper look at that transition. Sarah Werner suggests ways the traditional act of reading is forever altered by the nature of digital texts. From her experience in early modern book scholarship, she **builds a case** for how book history might inform our understanding of reading beyond the codex. Craig Mod speaks less of a one-way street from analog to digital and more of **a two-way street** between the virtual and the physical. As a designer of both websites and books, he brings to bear knowledge about the boundaries and content of ebooks and

physical books. Both sense a tension between physical and digital sources, methods, and productions.

Digital productions have many advantages, of course. They can help us to visualize and better understand the past, as Matthew Booker does by revealing the **physical and human alterations to San Francisco Bay** over the past two hundred years. Digital methods of communication also offer important opportunities to distribute our scholarly work more widely. Social media activity, Melissa Terras demonstrates, can **aggregate attention** to digital publications and projects available on the open web.

In a special section we showcase three recently-released projects which represent years of collaborative effort: **ORBIS: The Stanford Geospatial Network Model of the Roman World**, the **French Book Trade in Enlightenment Europe**, and **Mapping Texts**. These projects provide interfaces for large data sets of written documents and information about the physical world in order to enrich the study of ancient Rome, eighteenth-century Europe, and the modern United States. At the same time, they also enhance the questions scholars can ask of their sources, their technology, and their fields.

Although they have different topics and goals, each project blurs the categories of archive, tool, and publication. At the most basic level, ORBIS offers new methods for modeling extremely complex data; the French Book Trade in Enlightenment Europe introduces an interactive resource that will prove critical to disparate areas of the humanities; and Mapping Texts provides the opportunity for, and proves the necessity of, evaluating data that we already have.

The creators of each project introduce their intellectual goals and resulting design in an overview and a case study. Completing the issue are three independent reviews solicited by the *Journal of Digital*

*Humanities* that evaluate the topics and methods of these projects. As both the creators and reviewers note, these projects allow scholars to address questions arising from traditional concerns through the ability to hone in on very specific details and to uncover and visualize relationships. Ultimately, the deliberate design of the projects, in addition their sheer size, encourage new and unanticipated questions that may now be possible to answer.

Daniel J. Cohen and Joan Fragaszy Troyano, Editors

---

# Articles

## **Where Material Book Culture Meets Digital Humanities**

Sarah Werner . . . . . 2

## **The Digital-Physical: On Building Flipboard for iPhone and Finding Edges for Our Digital Narratives**

Craig Mod . . . . . 10

## **Visualizing San Francisco Bay's Forgotten Past**

Matthew M. Booker . . . . . 23

## **The Impact of Social Media on the Dissemination of Research: Results of an Experiment**

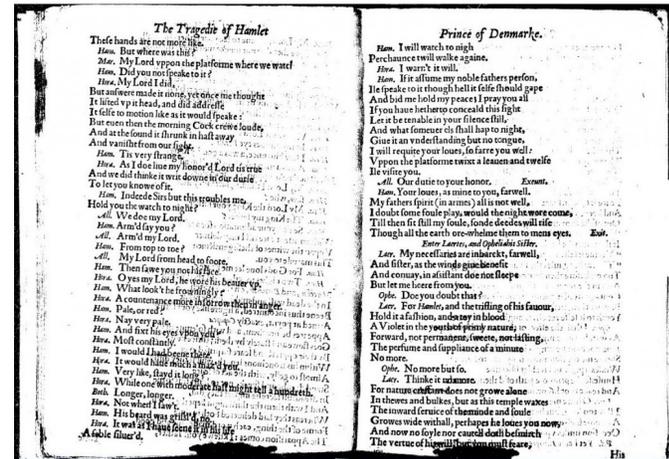
Melissa Terras . . . . . 30

SARAH WERNER

# Where Where Material Book Culture Meets Digital Humanities Meets Digital Humanities

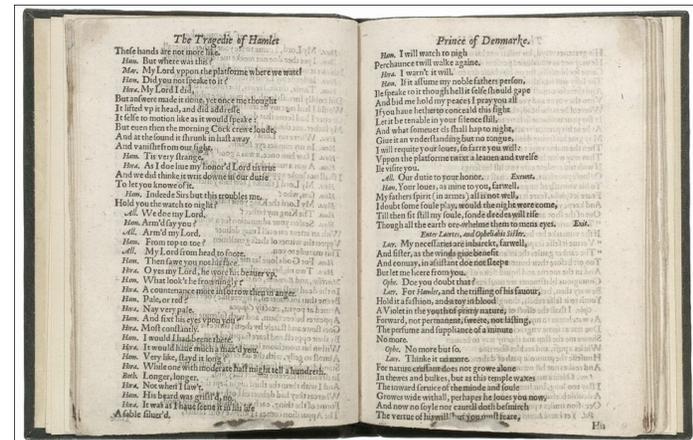
Discussions about early modern books and digital tools have tended to focus on one of two responses. One of the first things that people focus on is the amazing access to early modern works that digital tools have given us. Instead of schlepping from library to library across the globe — a series of journeys that many scholars can not easily afford — we can access nearly all extant early modern printed English books, and many continental ones, from our desktops. Thanks to digital collections like [EEBO](#) (Early English Books Online) and [ECCO](#) (Eighteenth Century Collections Online) and to catalogs like the [USTC](#) (Universal Short Title Catalogue, which links to images from many European collections), digital facsimiles are available for us to consult and download entire works from the early modern printed world.

There are limitations, of course. One is the quality of the images. EEBO consists of digital facsimiles not of early books, but of microfilms of early books. As a result, it doesn't always capture what we might want it to.



Opening From a Folger Q2 Hamlet, as in EEBO

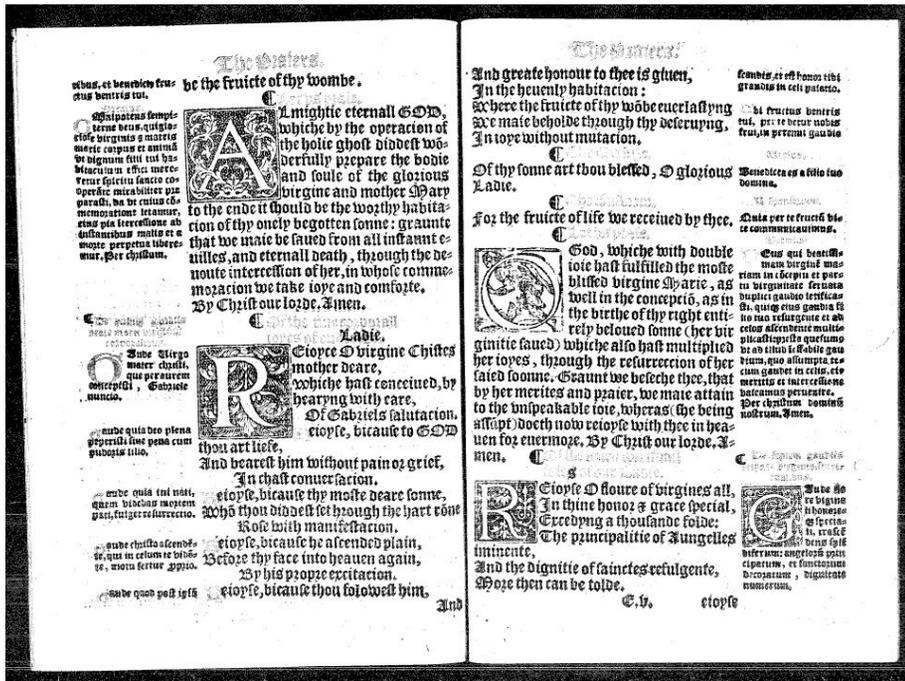
Above is an image from EEBO of the second quarto of *Hamlet* (STC 22276). You can see one column of text on each page, along with a whole bunch of other junk. Here's the same page opening from the Folger's reproduction of that book:



Same Opening, Same Copy, in a High Resolution Image  
(Click Caption to Open the Full Play)  
Image Used by Permission of the Folger Shakespeare Library

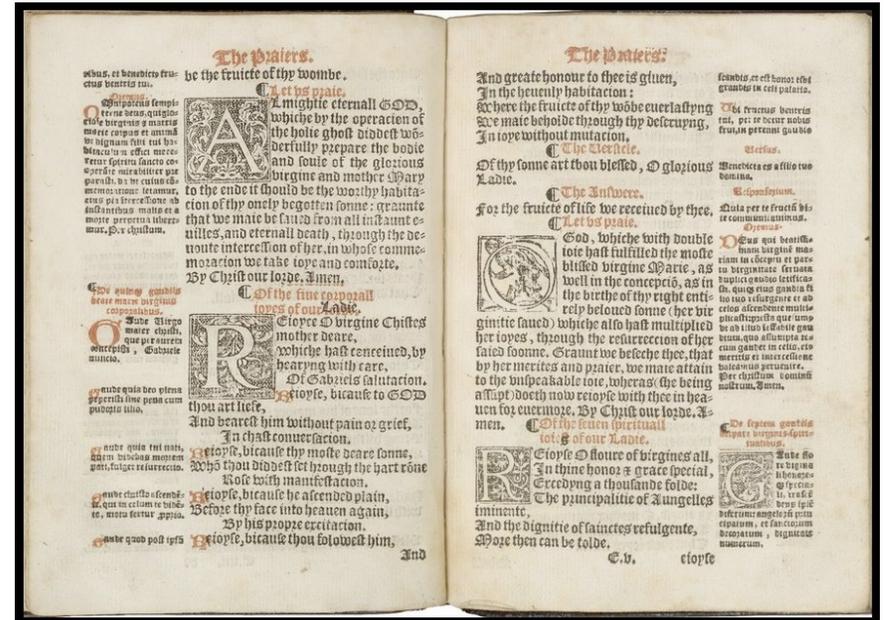
There's still ink bleeding through from the other sides of these leaves, but it's a bit easier to sort out what's what.

Or compare the reproduction of ink in this pair of images:

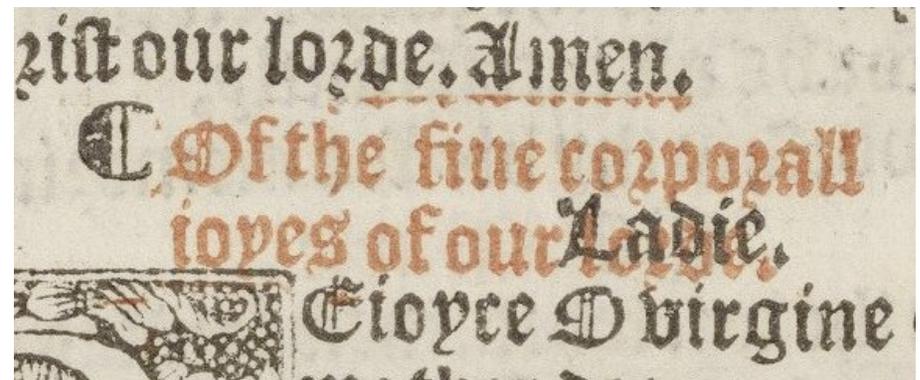


Opening From a 1557 Primer, as in EEBO

Because the red ink isn't visible in EEBO, you miss the dynamics between how the different categories of text play out on the page and what they might signal about the book's use. You also miss what's really a great mistake, the moment where the phrase "of the five corporall joyes of our lordie" is really a correction for the mistaken "joyes of our lorde."

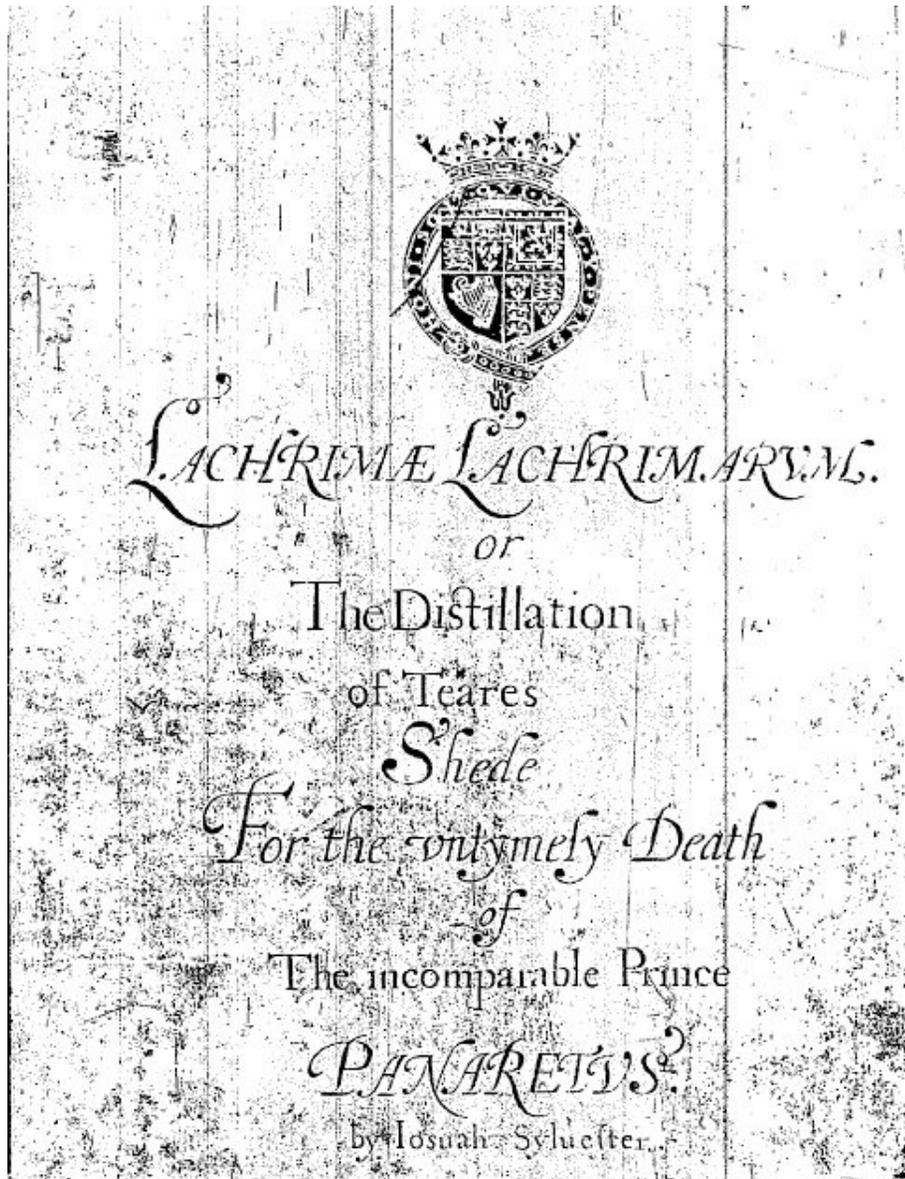


Same Opening, in a High-Resolution Image from the Folger (Click Caption for Zoomable Image)  
Image Used by Permission of the Folger Shakespeare Library



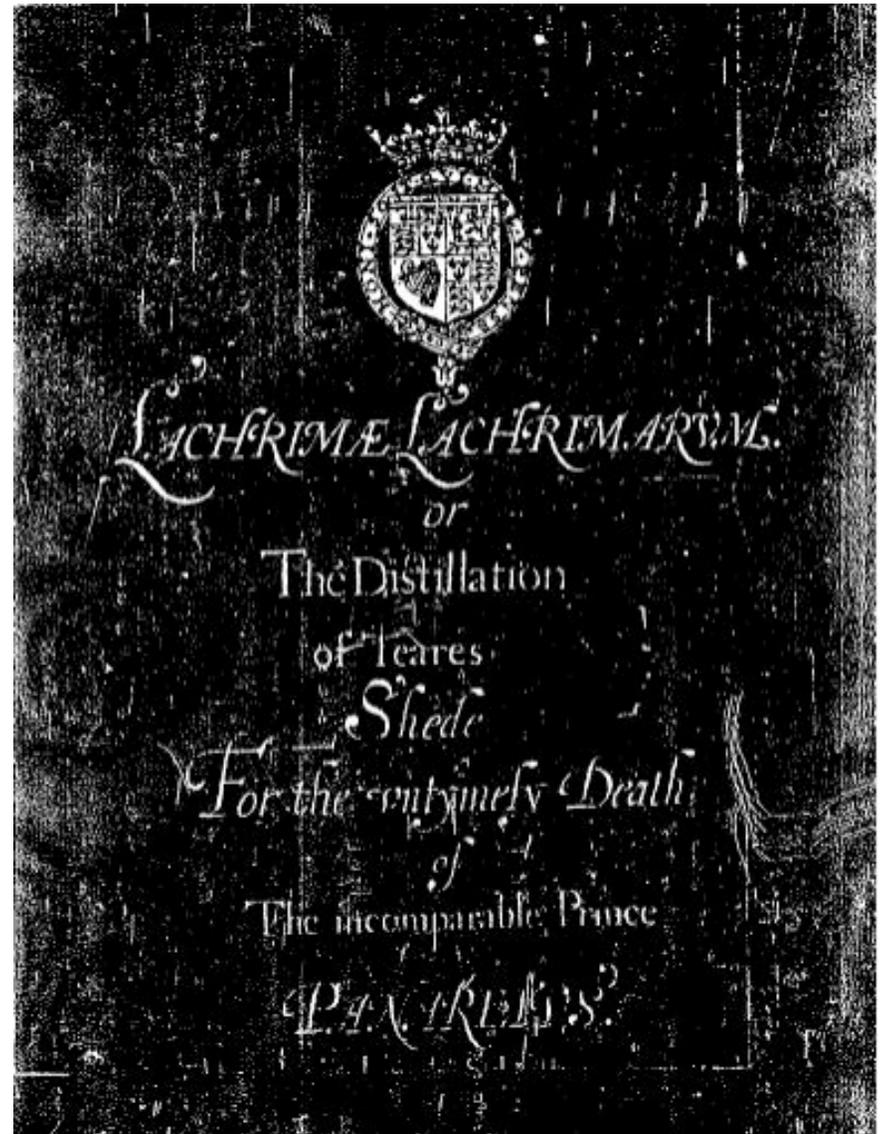
Never Mix Up Your Lord and Your Lady

My favorite EEBO moment, however, is this one: the title page of a 1612 elegy mourning the death of Prince Henry. This is how the image appears in EEBO . . .



Title Page of STC 23576 as in EEBO

. . . but this is how the image appears in their reproduction of the second state of this edition:



Title Page of STC 23577 as in EEBO

*Lachrimæ Lachrimarum* is a mourning book, and it was [printed](#) on pages bordered in black and sometimes entirely in black, with a xylographic title page (a title page in which white lettering appears on a black background). But when the microfilm was being processed, someone clearly didn't believe what they were seeing and they assumed it was a mistake, that it should be black on white.[1] And so they reversed the negative, producing a facsimile of a book that doesn't exist.

I have been complaining about the quality of reproductions here, and cost is a limitation as well.[2] But digital tools have, without a doubt, increased our access to facsimiles of early modern books. If I can sit in my study outside of Washington, D.C. and study [Erasmus's 1516 translation of the New Testament](#) by looking at a copy currently held in Basel, that's a win.

If one dominant way of thinking about digital tools and early modern books is in terms of access, another has been in terms of text. Access is about text, of course — what we're gaining access to is the ability to read texts. But there are also digital tools that don't simply read texts, they distant read them. [EEBO-TCP](#) can make research a bit easier if you're interested, say, in sassafras and want to find instances of it being discussed. In the right hands, you can do much more interesting types of computational analysis that can reveal things that would be difficult to see otherwise. [Recent work by Michael Witmore and Jonathan Hope](#), for instance, reveals that genre is marked not only in terms of plot, but also linguistically at the sentence level — histories and comedies and tragedies are genres that are grammatically inflected. That is surely a win, too.

The tools that I've just described rely on the ways we have always read books, albeit with increases in distance or speed (you can read a book

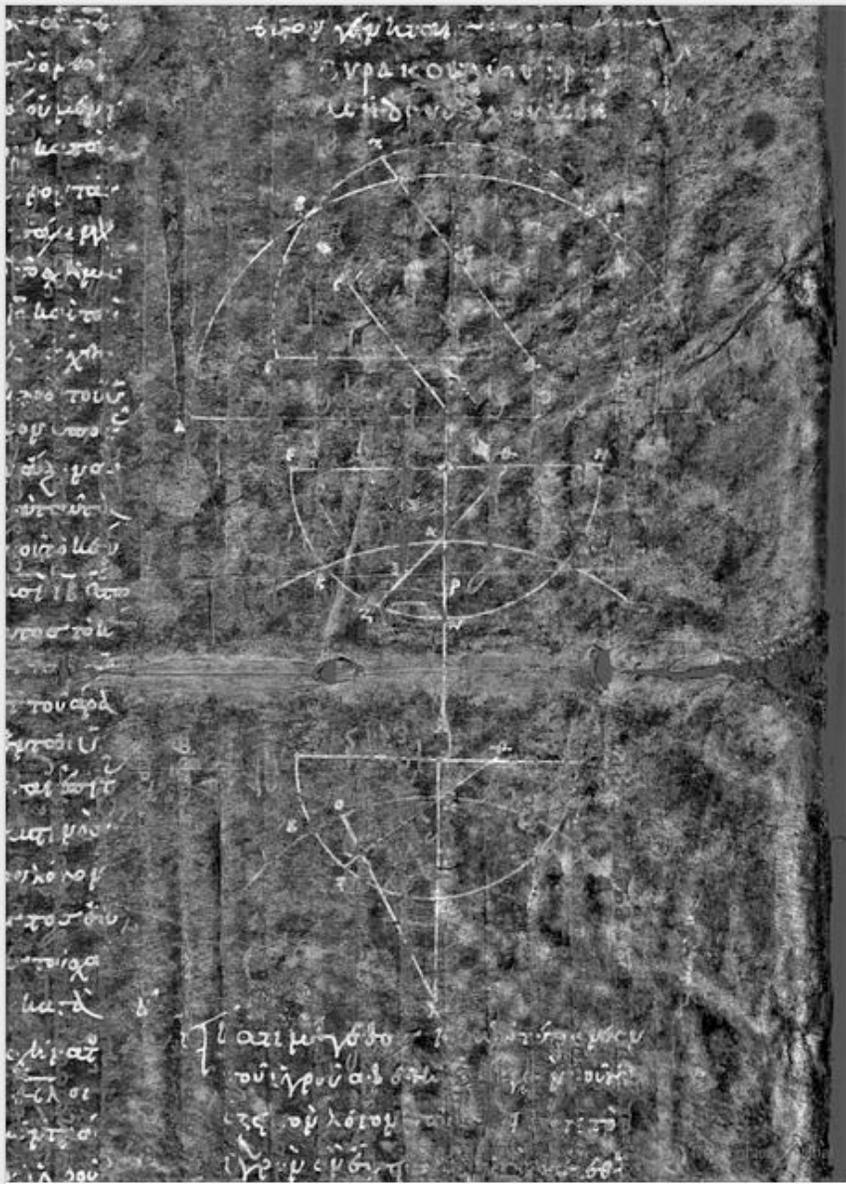
held at the Folger without leaving your study in Gdansk; you can analyze the texts of the entire Shakespeare corpus in a matter of minutes rather than years). But I want to wonder what new possibilities we might imagine. How might we use digital tools to look at texts differently? How might we use digital tools to represent texts differently? Can we move away from reading text to studying the physical characteristics of text, characteristics that can reveal important information about the content of the text and the cultural and historical creation of the artifact?



A Digitally Enhanced Image from Archimedes's Palimpsest (Copyright: the Owner of the Archimedes Palimpsest. Licensed for use under Creative Commons Attribution 3.0 Unported Access Rights.)

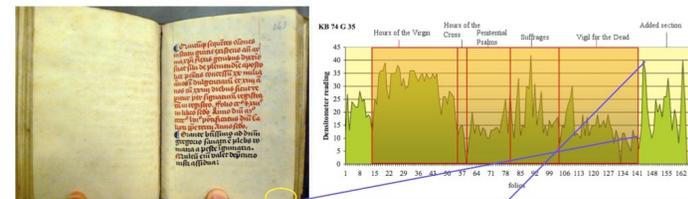
The multi-spectral imaging done by the Lazarus team of the [Archimedes Palimpsest](#) gives a hint of how digital tools might let us see things that would otherwise go unseen. The Archimedes Palimpsest is a 13th-century Byzantine prayerbook written over a 10th-century manuscript containing writings of the Greek mathematician Archimedes, as well as multiple other works from various periods. Using multi-spectral imaging, along with other tools, the team was able to recover visual access to much of the earliest writings in the book. Google took the project's dataset and made a ["Google book"](#) of

[the earliest state of the codex](#) resulting in a digital reproduction of a book that exists, but is not visible to us just by looking at it.



Cover of the Google Book of the Archimedes Palimpsest

One recent paper about [the use of densitometers to study levels of dirt on the pages of medieval manuscripts](#) suggests that we can learn about book usage through analyzing how and where dirt is distributed across a book. It might seem obvious that pages that are used more often will be dirtier, and that is in part what the author found, but the use of the densitometer revealed that it's more complicated than we can always assess with the naked eye. The paper's author, Kathryn Rudy, points out, for example, that she had assumed that two different patterns of dirt on an opening came from two different users, but the densitometer's analysis suggested that the patterns were similar enough that they were likely to have been made by the same person using two fingers to touch some pages rather than one. The analysis also pointed out that even books that retain visible marks might have been cleaned by modern owners to such a degree that the dirt is no longer viable as an analytical tool, something that might help us think about the changes books undergo during modern ownership.[3]



Fol. 142v-143r: Rubric with announcement of indulgence



Fol. 143v-144r: Incipit of the *Adoro te*



Fol. 144v-145r: *Adoro te*, cont'd

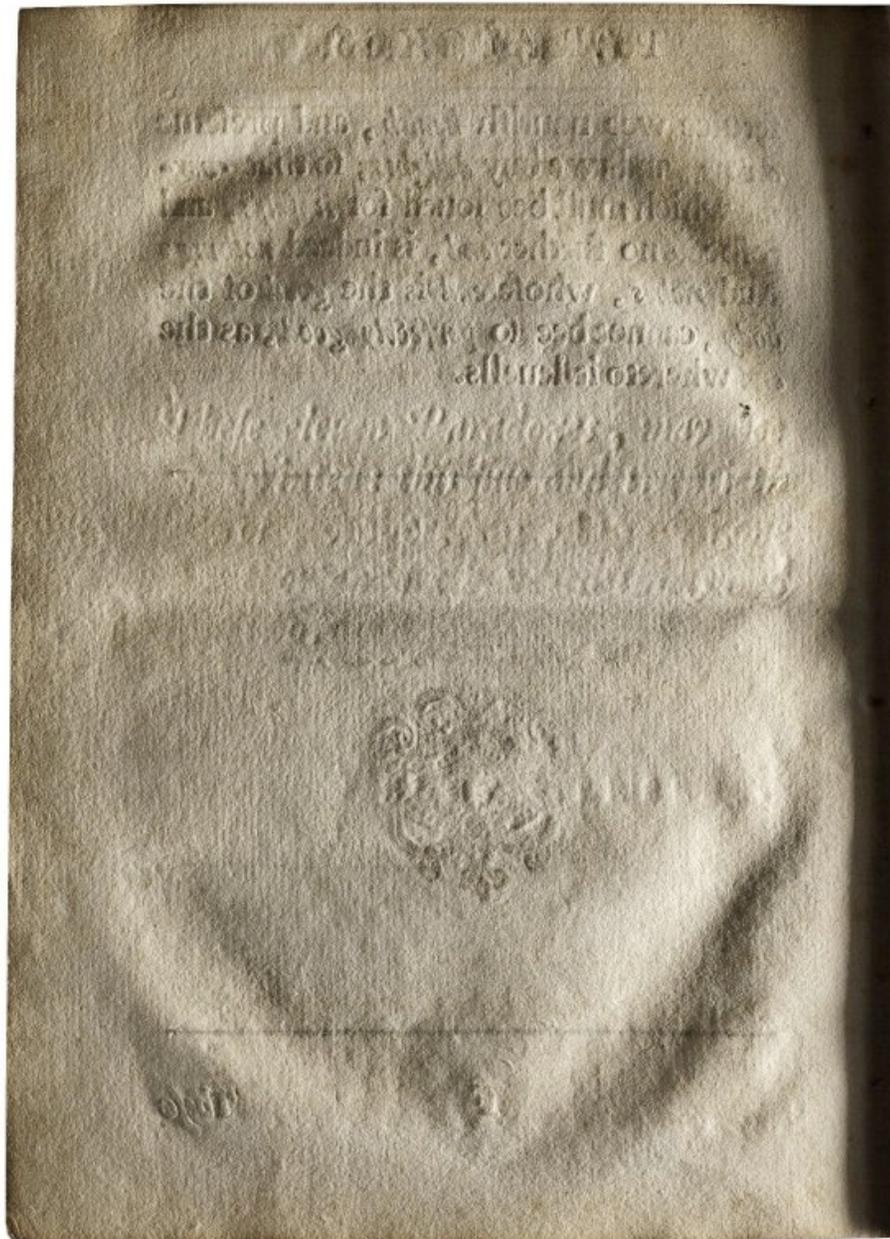
A Comparison of the Levels of Dirt Left on Leaves of a 15th-Century Book of Hours (Image Courtesy of Kathryn Rudy)

Studying the distribution of dirt is just the beginning of how we might begin to use technology to help us understand books in new ways. A colleague in Antwerp reports that German books held in Belgium smell different than German books held in Germany. The cause lies in how the paper was treated: paper needs to be treated with sizing agents so that it handles ink properly (instead of absorbing ink, ink sits on the surface of the paper and dries there, producing crisp and legible marks). His speculation is that books in Germany were sized in a multi-stage sequence, with the last step taking place after the book had been printed, perhaps as part of the binding process. Books that remained in Germany after they were printed went through this final process; books that were shipped outside of Germany seem to have missed that final stage, resulting in a noticeably different smell because of their different chemical properties. If this is the case, the smell of early German books can help scholars understand not only the physical acts of making paper and books, but can help us trace the circulation of early printed works. Using computers to analyze the smells of books and software to map those smells could help researchers learn how books were made and sold and used.

We could also use new technologies to explore the other senses we use when handling books. The feel of paper (or parchment) is another element of books that has more to offer than nostalgic fetishizing: the thickness, color, and pliability of paper can tell us about the costs of production, in part, but also give insight into the experience of using the book and its intended audience. How might the characteristics of feel be represented in digital media? Could a 3D printer replicate samples of different paper qualities? Could we project back from a paper's physical characteristics today to how it might have appeared and felt when it was made?

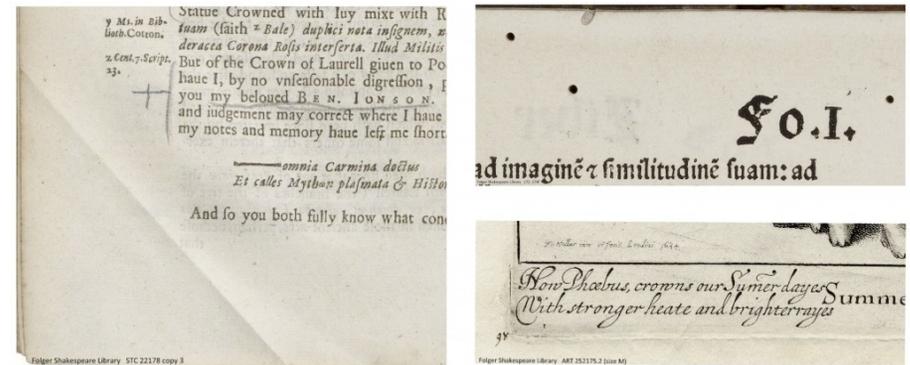
The three-dimensional aspects of paper extend beyond what can be felt by human touch. The process of making books in the letterpress period was a process of putting pressure on the paper, leaving behind an indentation on one side of the leaf and an extrusion on the other side of the leaf. In most cases, the indentations are visible because the instrument causing them (type, woodblock, stylus) left behind ink markings. In other cases, there are indentations without ink, sometimes caused when two sheets of paper are accidentally run through the press, sometimes left behind when the bearing type used to even out the blank spaces in a page leaves behind blind impressions.

There are also the indentations left behind during the papermaking process from the wires and frames used in the forms. Once we start thinking in these terms, we can find more topographical variations on leaves of paper: wormholes, dog-eared corners, holes left from stitches sewing gatherings and the binding together, plate marks from engravings. What might we learn from visualizing books not as texts to be read but as topographical maps?<sup>[4]</sup> The project to conserve The Great Parchment Book, a 1639 volume of records that was heavily damaged in a 1786 fire, works from the opposite perspective, [using digital tools to virtually flatten](#) a badly warped manuscript that cannot otherwise be read.



Folger STC 7043.2, Leaf F1v Under Raking Light  
 (Click Caption to See a Zoomable Version of this Image  
 Compared to One Under Normal Light)  
 Image Used by Permission of the Folger Shakespeare Library

Another option would be to use digital tools to visualize the context of books, to encounter them not in isolated codices, but in libraries. This [360° panoramic view inside the Strahov Monastery's Library in Prague](#) lets you see not only the entire room, but to zoom in to see the titles of the books on the shelves.



Folded Edges, Worm Holes, Plate Marks

This example [on the following page] is primarily a pretty picture, but imagine if this technology was married to something that let you look at catalog records of the books that you're seeing, or to switch from catalog records to a view of a book on a shelf. Could we rearrange the books on the shelves, grouping them by accession order, or the chronological order of their bindings, or by their provenance?

If we could use digital tools to estrange ourselves from our books, to defamiliarize what we think we know, we might learn something new about how they were made and how they are used. People keep pointing out to me that we are in the incunabula age of digital texts. We are. And replicating the familiar makes sense in the cradle days of digital texts. But let's not limit ourselves to reading the digital in the same ways we've always been reading.



Image by Jeffrey Martin  
[www.360cities.net](http://www.360cities.net)

This paper was given first as a talk at the University of Maryland "Geographies of Desire" Conference on April 28, 2012. It was originally published by Sarah Werner on [April 29, 2012](#) and was revised for the *Journal of Digital Humanities* in August 2012.

#### Notes:

- [1] Ian Gadd first pointed out this example to me.
- [2] There are resources that provide higher quality digital facsimiles of early modern books and that, unlike EEBO and ECCO, are free to

use. The [Folger Shakespeare Library](#) has digitized many works in their entirety, including all copies of the pre-1642 Shakespeare quartos and a couple of first folios. The [British Library](#) has digitized some of their collection, cover-to-cover, as have many other libraries, including that of the [University of Pennsylvania](#), [Princeton University](#), [University of Oklahoma](#), and the [Bavarian State Library](#). The [English Broadside Ballad Archive](#) now includes some high-resolution color facsimiles, and the [Universal Short-Title Catalogue](#) (covering all books printed in Europe in the 15th and 16th centuries) includes links to digital copies from many European libraries.

[3] Kathryn M. Rudy, "Dirty Books: Quantifying Patterns of Use in Medieval Manuscripts Using a Densitometer," *Journal of Historians of Netherlandish Art* 2:1-2 (2010). <http://www.jhna.org/index.php/past-issues/volume-2-issue-1-2/129-dirty-books>.

[4] Randall McLeod's work has helped me see the possibilities of thinking about printed books as textured objects. See, in particular, R. MacGeddon, "hammered" in *Negotiating the Jacobean Printed Book*, ed. Pete Langman (Farnham, Surrey, England; Burlington, VT, USA: Ashgate, 2011), 137-199.

CRAIG MOD

# The Digital-Physical: On Building Flipboard for iPhone and Finding Edges for Our Digital Narratives

## The Weight

We had been on a long journey but it was lost in the bits.

There's a stage in a product cycle where you *know* it's going to ship. Where you can see the end. It's *right there*, sitting at the corner of Emerson St. and University Ave. Or maybe sipping coffee at Fraiche.

*Oh, hello*, it waves — there. In front of you. The End. (*Or, An End.*) And seeing this puts you in a special space where when you think about it — think about all the work that it took the team to get there, to bring that end so close — you are overwhelmed with a flood of emotions.

It was November 2011 when we arrived at that point. Soon there would be an app in the Apple iOS App Store — something representing the top slice of tremendous work. But that version 1.0 ... how long would it last? How long before version 1.1 and 1.2? 2.0? Before we couldn't even remember how 1.0 worked or what the path there looked like?

I became curious. What *had* we created?

Poking around, I found:

*997 design comps in a shared folder,  
9,695 git commits,  
a bundle of notebooks full of sketches,  
and dozens of photographs from launch night.*

And so I asked myself a simple question: *What does that weigh?*

## The Bits

*Eight pounds.*

What?

*Eight pounds.* That's what it weighs.

But we'll get to that in a minute.

## On / Off

In James Gleick's [The Information](#), there's a moment where the very idea of information is abstracted to on and off states. Bits:

The bit is a fundamental particle of a different sort: not just tiny but abstract — a binary digit, a flip-flop, a yes-or-no.

And then, applied:

[Licklider] was working on an idea for quantizing speech — taking speech waves and reducing them to the smallest quantities that could be reproduced by a "flip-flop."

We've entered a similar binary on/off era for physicality. *Big* physicality. Star Trek style. To go digital-physical and back again is increasingly frictionless.

And so:

*What do we gain from these jumps?*

*How can they reframe experiences to help us better understand them?*

These are questions I've found myself returning to repeatedly these past few years.

Abstractly, you can think about going from digital to physical as going from *boundless* to *bounded*. A space without implicit *edges* to one composed entirely of edges.

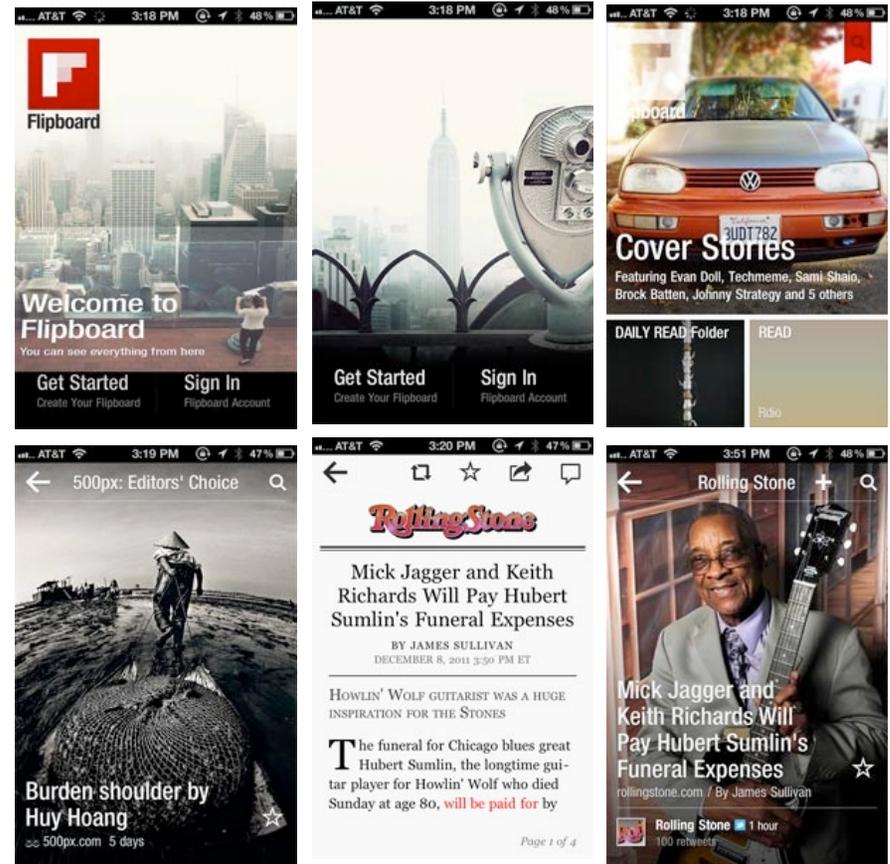
For a while it had been clear to all of us that edges are a critical framing aid in helping us *consume*<sup>[1]</sup> but it wasn't until last year — helping build Flipboard for iPhone — that I began to understand how critical they are to gain perspective on *creation*. To gain perspective on a journey captured in bits.

This is an essay about recognizing and reorganizing our journeys that live largely in digital space. How do we ground and bind those experiences? What is the value in giving them edges so we may hold them in our hands and hope to understand, perhaps, the weight of the work we produce?

## The App

**December 2011**

The team had spent a great deal of the year crafting, refining, building, and rebuilding [Flipboard for iPhone](#). It was a demanding journey during which a prodigious number of ideas were challenged. Time and time again, the team reconsidered a button here, a transition (*or two*) there.



### FLIPBOARD FOR IPHONE V1

By December, every piece of the application had been disassembled, scrubbed, and oiled from a user experience, surface design, and information architecture perspective. Furthermore, every piece of the underlying engineering had been hammered and abused in testing for solidity: just try to crash the thing.

I mention all of this to emphasize the great experiential *texture* we felt during the process of building this application. Clearly, we were making *stuff*. Lots of it. Making and throwing away and making more

stuff. Building upon and learning from the ever growing pile of experiments. Throughout, the team iterated in parallel on several design and engineering loops across everything. The final application is a piece of genuine craftsmanship produced by a full-stack team of genuine craftsmen.

This layered process happened almost entirely in digital space. Design comps were produced in Photoshop or Fireworks, screens mirrored to iPhones. Folders were shared throughout the team. And the IA was specced in Illustrator or InDesign.

Even the iOS software changes were captured with an atomic granularity. In most contemporary development environments, when an engineer modifies a program, that modification is checked back into the main source code repository along with what's called a "commit message." In this message is a brief description of the changes made by that programmer. In big projects, there can easily be thousands of commits. Some may be tiny: "Changed transition speed to 0.6 from 0.4 seconds." Others, much longer or more affecting: "Switching to live servers!" The codebase for Flipboard for iPhone is composed of nearly 10,000 such commits.[2]

And so we built. Commits were committed, design folders filled up, and screenshots cluttered up the photo folders on our devices. In other words, the more we iterated the more *digital detritus* built up.

### ***Digitally Thin***

There's a feeling of thinness that I believe many of us grapple with working digitally. It's a product of the ethereality inherent to computer work. The more the entirety of the creation process lives in bits, the less solid the things we're creating feel in our minds.[3] Put in more concrete terms: a folder with one item looks just like a folder with a

billion items. *Feels* just like a folder with a billion items. And even then, when open, with most of our current interfaces, we see at best only a screenful of information, a handful of items at a time.

Perceptually, beyond some low threshold, data becomes boundless to us. Cloud storage compounds this: we don't even worry about hard drives filling up anymore! Even when digital streams have clear beginnings and ends, I think we — humans — do a bad job at keeping those edges in view. In trying to reflect upon vast experiences or datasets captured entirely in bits with most standard interfaces, we run into the same wall as in trying to imagine infinity: we can't.[4]

### ***Finishing iPhone***

As Flipboard for iPhone was nearing completion, I began to think about this detritus — our narrative; the *proof* of our journey. What struck me is that despite knowing we had been on a long journey, it didn't feel like that journey was manifest anywhere.

Sure, you could open the design folder and cover flow through our thousand design comps. You could peek in the git repository and scroll through the near infinite number of commit messages. But, still: that thinness! The experiential texture of the journey was butting against the singularity — that fog of immateriality — that information enters when made digital.

I was leaving the company at the end of the year and I needed something to *represent* that journey. To give it edges, for me. For the company. So I did what I do — I flip-flopped the data. I made a book.



## The Umbrellas

But let's digress for a second.

Books have always excelled at capturing process. But perhaps not the process we're talking about here. And this is an important distinction.

As the sun rose on October 9, 1991, Christo and Jeanne-Claude unveiled one of their temporary "gentle disturbances"<sup>[5]</sup> — The Umbrellas. Across portions of the eastern coast of Japan and the western coast of the US, they opened 3,100 massive umbrellas. In their words:

"This Japan-USA temporary work of art reflected the similarities and differences in the ways of life and the use of the land in two inland valleys, one 12 miles (19 kilometers) long in Japan, and the other 18 miles (29 kilometers) long in the USA."<sup>[6]</sup>

This was the result:



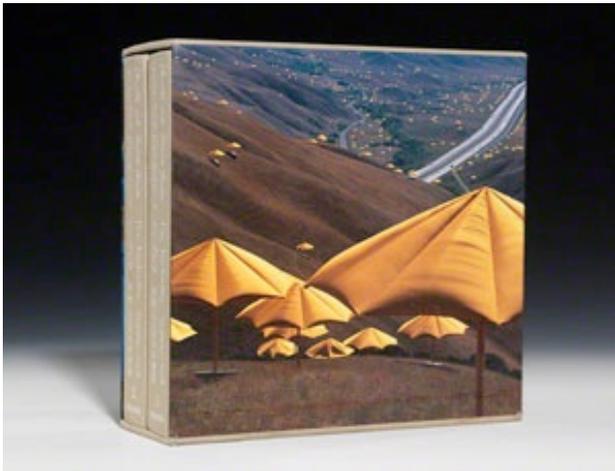
THE UMBRELLAS — CALIFORNIA SIDE

Wonderful stuff.

Years ago, standing in a small bookshop in London, I happened upon a large table far off in the back corner. Atop it were two hulking volumes. Together they comprised the Taschen collectors edition for The Umbrellas; I was mesmerized.

For hours I stood there plying through those two objects which so well captured Christo and Jeanne-Claude's creative process. Documented within was the residue of engineering a dreamscape and overcoming

endless bureaucratic snafus. That book — the physicality of it — codified the great effort behind their beautiful mountainside and rice paddy installations. It memorialized their efforts and journey. But for whom was this book created? I suppose it was for me — someone on the outside.



THE UMBRELLAS, THE BOOK(S)

The magnitude and grandiose nature of their work was present all around their home, their work space. It manifest physically in those files and papers and cabinets. For them, a monster book like this didn't illuminate the enormity of their undertaking — they were aware of The Umbrellas' bombast every time they opened their studio door.

It's also important to note that Christo and Jeanne-Claude didn't directly take money from outsiders. In fact, they funded much of their projects by selling the detritus of their process:

[The] 26 million dollar temporary work of art was entirely financed ... through the sale of the studies, preparatory drawings, collages, scale models, early works, and original lithographs.[7]

I would argue, however, that for Christo and Jeanne-Claude, the book simply formalized much of what they already knew. Their work benches were covered in schematics, file cabinets filled with correspondence between farmers and governments and architects and textile manufacturers and engineers. In other

words — to them the

Twenty-six million dollars of sketches! You can imagine just how viscerally aware they were of *process* by raising that kind of cash through selling off the very goop of their creativity.



PREPATORY SKETCHES:  
THE CALIFORNIA SIDE



PREPATORY SKETCHES:  
THE JAPAN SIDE

Their monograph gave form to their creative process for those of us not privy to studio access; to those of us who had never seen their chaotic lair of physical detritus.

So this book about The Umbrellas is for *us*. It captures their process and — if not entirely making sense of it — at least gives us a frame by which to view it. *Edges*. But for Christo and Jeanne-Claude, the book was simply a formality. If anything, it *shrank* their sense of the project from something full of physicality to just a couple of books.

Which brings us back to the previous point: Something curious happens to our ability to understand scope when we move all that goop of process and narrative into a computer. Even for those of us *doing* the making: the insiders. The Christos, if you will.

When all the correspondence, designing, thinking, sketching — the entirety of the creative process — happens in bits, we lose a connection. It's as if all that process is conceptually reduced to a single point — something weightless and unbounded. Compounded over time, the understanding of where one is as a creative in a digital landscape collapses to the just-a-little-while-ago, the now, and maybe the tomorrow.[8]

At the end of the Flipboard for iPhone project I wondered if we could find value by *exploding* that singular point into a form we could grasp. That form could have taken many shapes but it's easier than ever to make a book. Although, this would be a book to serve the precise opposite role of *The Umbrellas* monograph. It wouldn't shrink our creation. Instead, it would hopefully give shape and weight to the amorphous nature of our digital production processes.

## The Book

And so:

*997 design comps in a shared folder,  
9,695 git commits,  
a bundle of notebooks full of sketches,  
and dozens of photographs from launch night.*

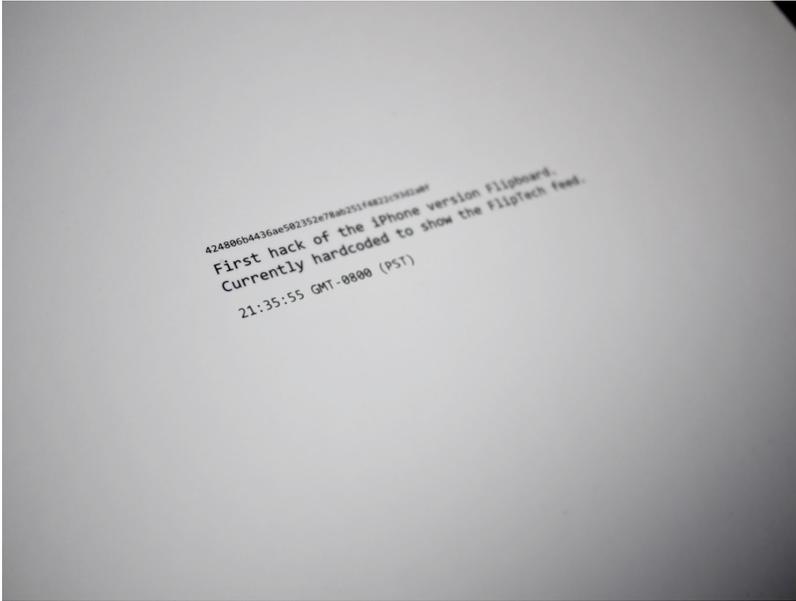
*What does that weigh?*

At 276 pages long and 1 foot by 1 foot square, it weighs nearly 8lbs.



1 FOOT BY 1 FOOT — PHOTO BY MARCOS WESKAMP

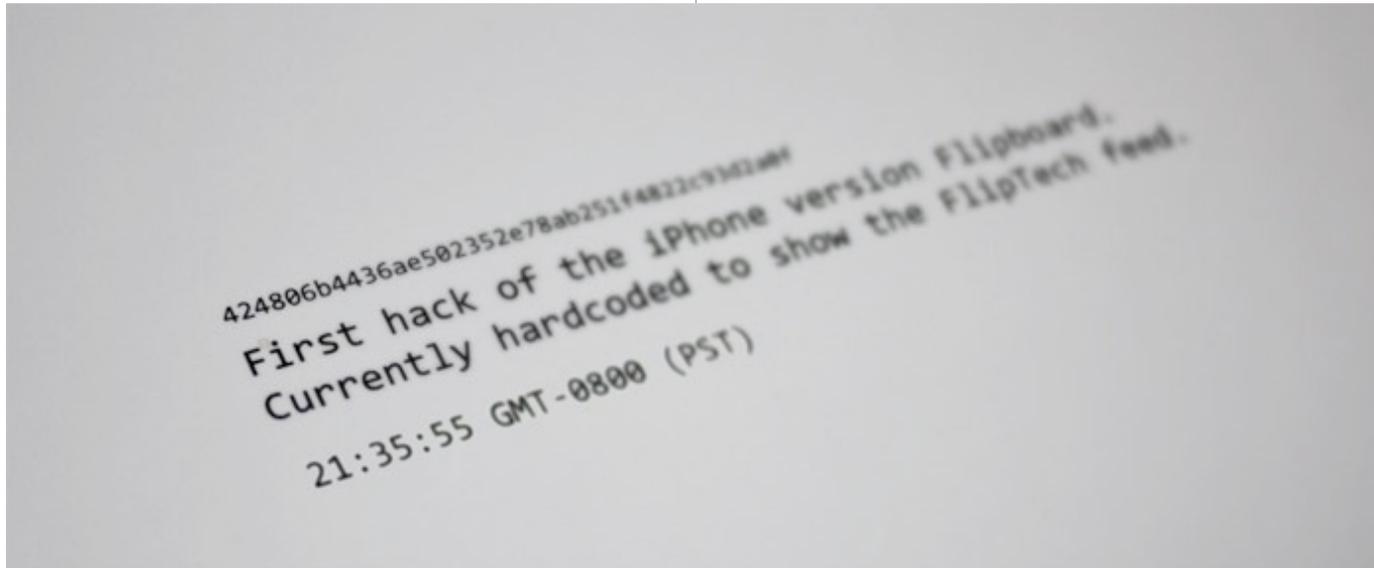
It begins on the cover with a commit message — the first push to the main code-base. A hack. A test, a possibility. The exact moment development begins on a software project is always fuzzy. Lots of development happened before this moment. Other intense cycles happened after. But this moment marks the first time in the code stack that the product had an identity — the seed for what was to come.



424806b4436ae502352e78ab251f48222c9302ae4  
First hack of the iPhone version Flipboard.  
Currently hardcoded to show the FlipTech feed.  
21:35:55 GMT-0800 (PST)



424806b4436ae502352e78ab251f48222c9302ae4  
First hack of the iPhone version Flipboard.  
Currently hardcoded to show the FlipTech feed.  
21:35:55 GMT-0800 (PST)



424806b4436ae502352e78ab251f48222c9302ae4  
First hack of the iPhone version Flipboard.  
Currently hardcoded to show the FlipTech feed.  
21:35:55 GMT-0800 (PST)

### FLIPBOARD FOR IPHONE THE BOOK — FIRST COMMIT, DETAIL

I love the hyper granularity of time — 9:35pm and 55 seconds on a Saturday

night on the west coast (GMT -0800) of the United States. So specific!

Romantics like to correlate the loss of physicality in our digital shifts with a certain loss of humanity. Often cited are the correspondences between Raymond Carver and his editor Gordon Lish in which we see clearly — thanks to the pencil strokes and marginalia of their found

Herb finished his drink. ~~Then he got slowly up from the table and said, "Excuse me. I'll go shower." He left the kitchen and walked slowly down the hall to the bathroom. He shut the door behind him.~~ "Gin's gone," Herb said.

### LISH'S CHANGES TO CARVER'S STORY BEGINNERS

letters[9] — just how instrumental Lish was in shaping Carver's voice.[10]

Those early edits[11] were so grandiose — pulling the text into such a minimalist place — that when Carver dropped Lish and used other editors later in his career, readers misattributed the shift in writing style:

... those who viewed Carver's later stories as more expansive than his early work, simply never knew that he had always been expansive.[12]

If we're excited by the voyeuristic insight of that, we should be *doubly* excited about where things are going.

This — this hyper-*hyper* granularity of time on the *Flipboard for iPhone* book cover — 9:35pm and 55 seconds; digital atomic precision of that moment — is just the tip of the commit-stack for what kinds of meta-data are now — or soon will be — generated within the digital texts we produce.

Perhaps the next Carver's manuscript will contain the entire typing history of the document including GPS data of where he was when he wrote it. We will be able to replay the entire composition process. Shadow, if you so desire, a particular Hemingway through a certain Spain as he writes a new *The Sun Also Rises*.

Here's that voyeurism made real: In 2011 the startup Stypi released a product[13] which already allows you to replay document creation. Every keystroke. By sharing the URL of a document, readers play back your thought process. YCombinator founder Paul Graham (*and Stypi investor*) used this application to compose his November 2011 essay, *Startups in 13 Sentences*. [14] It's fascinating and illuminating to watch Paul write. We gain an insight — a very human, very intimate one — into how he crafts his essays.



FLIPBOARD FOR IPHONE THE BOOK — FRONT COVER



FLIPBOARD FOR IPHONE THE BOOK — BACK COVER

So the Flipboard for iPhone book begins with some of this meta-specificity on the front cover and ends with the final commit message on the back. It uses the natural digital boundaries of version one of the app — first and final commits — to define the boundaries of the digital-physical version.

The first commit happened as a late Saturday night hack. The final commit happened on a Thursday at 4:47am after a marathon week of coding. There's something about this specificity that's perversely romantic. As if you can imagine the very moment the engineers hit the return key to finalize these decisions. Flipboard for iPhone was born precisely, and it was finished precisely.

The book then steps through, month-by-month, the design and engineering of the product. It begins in February with HTML comps, and ends in December with version one of the application.

What we're able to see in this form — printed — perhaps better than we can in cover flow, is the sheer tightness of our design iteration loops. Pages of information architecture comps, sketches from our notebooks, and views of the various grids through which we iterated take on new clarity in this digital to physical flip-flop.<sup>[15]</sup>

Then at the back: the commits. All 9,569 of them.<sup>[16]</sup> As you flip through this high-resolution macro view of the commit stack, the very texture of the page changes as new developers join the project; as the software jumps into different phases of development.

And finally, at the very end, pages and pages of the photos from our launch night. They were covertly assembled via geotagged searches across Instagram and Flickr.

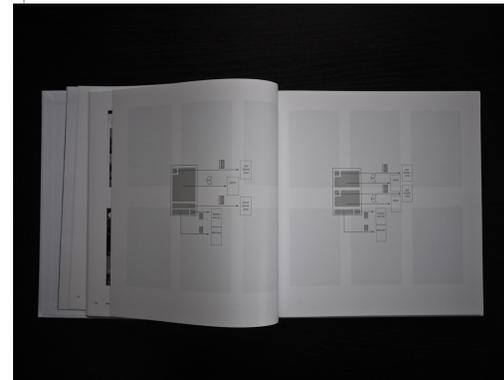
All the digital intangibility of iPhone development is inscribed into this eight pound book, wrapped in an immutability. It's a digital-physical. An explosion of months of ethereality into something human parseable.



FLIPBOARD FOR IPHONE THE BOOK — ITERATE ITERATE ITERATE



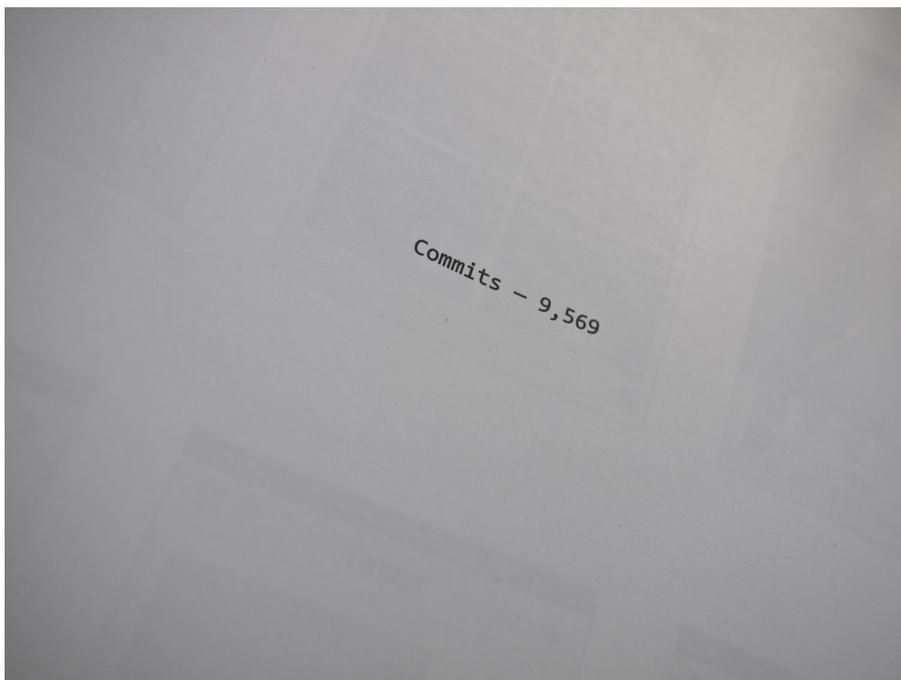
FLIPBOARD FOR IPHONE THE BOOK: ITERATE ITERATE ITERATE, MORE



FLIPBOARD FOR IPHONE THE BOOK — IA DEFINITION



FLIPBOARD FOR IPHONE THE BOOK — GRID IT UP



## The Delivery

### *UPS*

The book arrived from the print on demand company Blurb[17] without fanfare — a square box from UPS simply appeared on my desk one morning.

*So this is what all of that design and development weighs.* As I opened the box and pulled out the book, there was something deeply satisfying about its imposing physicality. An emotional response you couldn't intuit from the InDesign templates. This feeling, of course, speaks to the cognitive dissonance we experience when jumping between digital and physical representations of the same thing. Especially when they're composed of the same data.

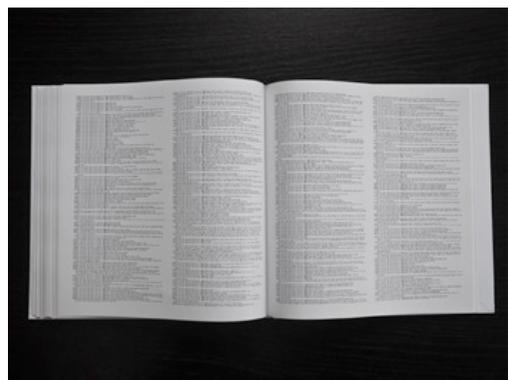
### *The Fellowship*

The post-launch project breakdown happened in a quiet, sunlit room in the early afternoon. The Flipboard for iPhone team gathered around a giant wooden table. Ten or fifteen of us. What went right? What went wrong? What could we do better next time?

We talked about what a wild journey we had been on. *But*, I piped up, I worried that the journey — and with it lessons we learned — seemed to be fading in the post-launch quickness that often characterizes Silicon Valley. What did we do right and wrong are great questions, but even more generally: What did we *do*?

It represented closure on a process that often has no closure.

I pulled from my bag the giant book and placed it on the table with a thud. And as the team began to flip through it and understand what it was, you could feel the room swell with a strange relief. Finally there were edges to this intensely immaterial process we had all been



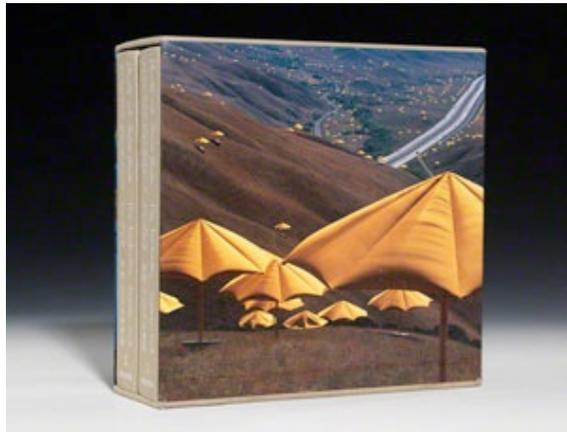
9,569 COMMITS



LAUNCH PHOTOS

through. The book represented, I believe, closure on a process that often has no closure. And pragmatically, it would now serve as a repository of the rich interface experiments we had conducted.

The book wasn't made for anyone outside of the company. It is, weirdly, an object for the folks who made the very thing it's about. It's for the people who should be most intimate with the project and process but who — because of the nature of digital — may have a hard time seeing the edges and understanding the weight of what they had created.



THE UMBRELLAS

*The Umbrellas* is for the consumers.

*Flipboard for iPhone* is for the producers.



FLIPBOARD FOR IPHONE

## YES/NO, ON/OFF

It's worth remembering that a hardcover book like this is as much a product of the digital publishing shift as any .epub or .mobi file. *Flipboard for iPhone* would have been nearly impossible to make five years ago. I produced two copies. Two copies! of this huge thing, printed in full color. And they arrived ten days after submitting the files. A miracle of sorts.

... a binary digit, a flip-flop, a yes-or-no.

Digital ↔ Physical.

### *In Part*

What could possibly represent that scale of emotion?

Of course, *Flipboard for iPhone* is not an accurate representation of the *full* production process. Anyone who has been involved with a startup can attest — imbued in a company's output is an incalculable amount of team energy. Lost sleep, lost muscle mass, stress, highs as a beautiful solution to an engineering or design problem emerges, lows when you realize the next solution obviates the first, bonds born from late night hacking sessions, design reviews, inside jokes, scaps, animated gifs, high-fives, scowls. All of this is manifest in the final product *somewhere*. It's in there. All of that and more. Of course you can never show it all. What could possibly represent that scale of emotion? A mountain? All you can hope to find is something — a scaffolding — to represent a clear subset of the process. A well defined space onto which the team can affix those experiences.

And so what projects like this speak to is the unique and increasingly important value we can give data by abstracting physicality. Jumping back and forth. Creating that space. Capturing a journey effortlessly in

bits, and then giving it *edges*. This dance makes our digital experiences more understandable, parseable, consumable.

Edges are about *feeling* as much as *seeing*. With edges comes a sense of weight. And with that comes the ability to feel — physically and psychically. And with that, a better understanding of what we've built and where we've been.

Originally published by Craig Mod at [craigmod.com](http://craigmod.com).

This work is distributed under the [Creative Commons Attribution-Noncommercial-Share Alike](https://creativecommons.org/licenses/by-nc-sa/4.0/) license. If you wish to use reproduce any of this content in a commercial context, explicit permission is required. Please contact the author directly.

#### Notes:

[1] It's one reason, for example, it feels so good to scroll down to most e-mailed on nytimes.com — it's bounded, doesn't update much, and you feel like you can read it all.

[2] If you want to geek out on this stuff, you can learn all about version control and software like [git](https://en.wikipedia.org/wiki/Revision_control) here: [http://en.wikipedia.org/wiki/Revision\\_control](http://en.wikipedia.org/wiki/Revision_control).

[3] I believe part of the recent surge in the desire to do "real" stuff with our "hands" is driven by the digital thinness we're wrapped in. Conferences like [Do Lectures](#), new shops selling [pickles in Brooklyn](#), [woody digital refuges](#). It feels like there's a renewed momentum around craft and physicality. All circumstantial, of course, and perhaps just a product of my generation getting sick of screens.

[4] Of course, this is mainly an *interface* problem. That is to say — printing stuff out isn't the only way to draw edges, combat the feeling of thinness, or help us keep digital data in perspective. Collating and printing is just one *kind* of interface with which to attack this problem. Higher resolution screens with smarter design solutions can and should also help us solve this problem.

[5] <http://www.artagogo.com/commentary/christo/christo.htm>.

[6] [http://www.christojeanneclaude.net/major\\_umbrellas.shtml](http://www.christojeanneclaude.net/major_umbrellas.shtml).

[7] [http://www.christojeanneclaude.net/major\\_umbrellas.shtml](http://www.christojeanneclaude.net/major_umbrellas.shtml).

[8] This is why great project managers are so critical — their job is to keep the entirety of the production process in view, both the long-before and the long-after our now. They are masters at framing all the intangibility of a software project into something with clear edges, and making sure all the pieces — front and back-end engineering, design, marketing, etc — are all moving at the right speed, at the right place within the context of this greater whole. Kind of like a rally car [co-driver](#).

[9] Excerpts from their letters. See Raymond Carver, [Letters to an Editor](#), *The New Yorker*, December 24, 2007.

[10] The full edit of the "*Beginners*" text is available on newyorker.com: ["Beginners," Edited: The transformation of a Raymond Carver classic.](#), *The New Yorker*, 2007.

[11] [Being Raymond Carver](#), *The New Yorker*.

[12] Motoko Rich, [The Real Carver: Expansive or Minimal?](#) *The New York Times*, October 17, 2007.

[13] Based off of [Etherpad](#) which was acquired by Google in 2009, open sourced and the original site shutdown. <http://en.wikipedia.org/wiki/Etherpad>.

[14] Paul originally linked to the article on [his blog](#). You can see the original 'performance' on [Stypi](#) (nice URLs!). And read the reactions to the piece here on Graham's own [Hacker News](#).

[15] Again — one of the qualities we're benefitting from in printing these out is resolution and size. If we have paper-resolution screens (like the 2012 iPad) at a larger size, then I think we can achieve *aspects* of the experience we get from printing it out on paper.

[16] `git log --pretty=format:'%h %an %ai %s' {{{first_commit_hash}}}..HEAD --reverse > everything.txt :: More on hacking git logs.`

[17] I've been using [Blurb](#) to do print on demand books for years now. They provide InDesign templates on their site as well as PDF specifications. They offer branding free (for a fee) books with a variety of great paper and binding options. Also, as a bonus, Blurb is a great company and Eileen Gittins, their CEO, is wholly devoted to making beautiful books accessible.

MATTHEW M. BOOKER

---

# Visualizing San Francisco Bay's Forgotten Past

To begin: One question and one metaphor. The question is, what use is visualization to historians? How does this method add value to the work we do? The metaphor is accretion, the term geologists use to describe the building up of new soils through deposits of materials eroded elsewhere. Accretion works to describe the process by which generations have modified San Francisco Bay without destroying that which came before. That history remains, even if it is invisible to the naked eye. Visualization helps us recapture the forgotten past.

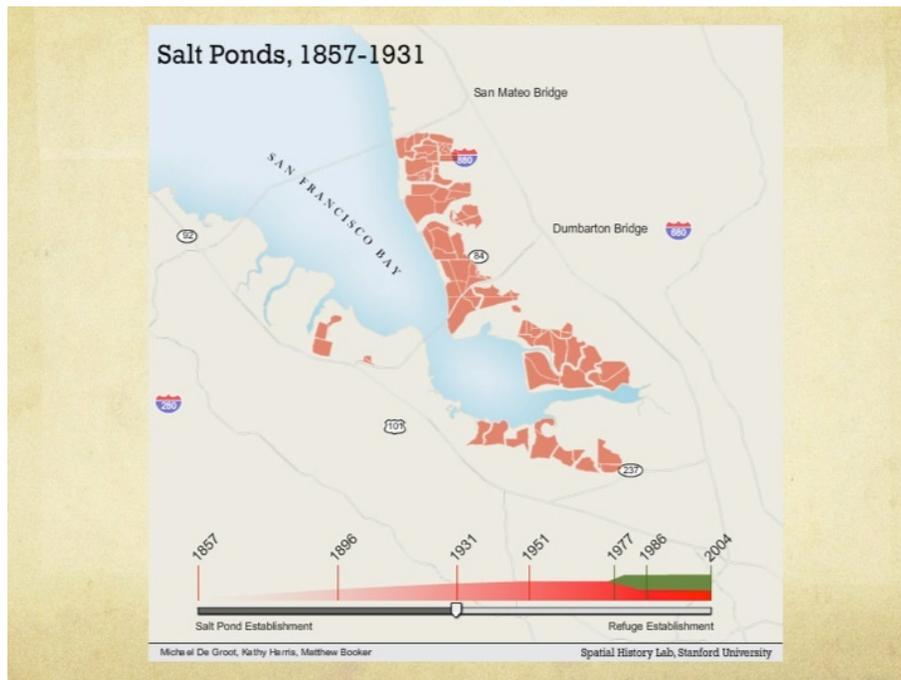
Archaeologists see the past as a series of layers. They dig through one past into another. Historians can think in similar terms. Many people in the American West think of nature as timeless and human creations as quite recent, even superficial. The San Francisco Bay is in fact a young land with deep human history. The current bay has only existed for some five thousand years or so, as rising seas flooded river valleys already inhabited by Indian peoples. Their forgotten but not erased past was visible in the hundreds of shellmounds these people raised over the millennia.

But rather than this astonishing story of Indian persistence — adapting to environmental change to inhabit the same places for more than a thousand years! — many Americans think San Francisco Bay's history began with one of the great geo-engineering events in American history, the California Gold Rush.

The usual story of the Gold Rush is the story of individual miners making their fortune. In fact most miners never did find wealth. Nor was this a story about individuals. Surface gold quickly ran out and the miners with pans gave way to the nation's most heavily capitalized corporations. Mining companies hired men to dam rivers and use water cannons to wash down the mountains for the gold inside. From the 1860s through the 1880s, hydraulic miners washed more soil, sand, and gravel out of California's Sierra Nevada Mountains than all of the earth ever excavated from the Panama Canal. This soil filled parts of San Francisco Bay by more than a meter. The gold rush also spurred another, lesser known mining industry, one that flowed in the opposite direction.

Beginning in the 1850s, salt miners converted much of south San Francisco Bay's tidal marsh and wetlands into evaporative salt ponds. Salt miners adapted an ancient method to industrial production. The ancient part was the steady ocean wind evaporating water and leaving salt behind. The industrial part was the use of windmills, later electric pumps and bulldozers to move the brine through a series of pools and to produce millions of tons of sea salt each year.

Salt was and remains one of the most important mining industries in California. Indeed by volume and by value salt often surpassed gold. By the 1930s salt mining was among the largest land uses in the San Francisco Bay region.



South San Francisco Bay Salt Ponds, 1857-1931

Only a tiny fraction of this salt went to preserve food, such as in the canneries of the Sacramento Delta where generations of Asian and later Mexican workers packed asparagus for the world market. Most salt went to mines in the mountains where it leached gold and silver from rock, to chemical plants to make bleaching agents for the paper mills of the Pacific Northwest, to Western oil refineries to make gasoline, to water treatment plants for the West's burgeoning cities, and to make napalm jelly for the war in Vietnam.

Here too what began as a family business rapidly became a corporate endeavor, then a monopoly as one company, Leslie Salt, bought up all the others and converted almost the entire shoreline of the bay south of San Francisco into vast stagnant pools walled off from the tides of the bay.



Leslie Salt Company Ponds, 1931-1970s

When salt companies converted the marshes and beaches of the bay into salt ponds they destroyed other uses of the edge. During the nineteenth century, the margin of San Francisco Bay was a public space, a place where people picnicked, played, rested, made love — and where thousands of people hunted, fished, and foraged for food. These uses depended on the bay remaining a commons, a shared space open to all. The mudflats and marshes were not just ideal locations to build salt mines. They were also the bay's nursery for young fish and crabs, the most important links in the Pacific flyway hosting millions of ducks and geese each year, and vast factories for shellfish and shrimp.

In fact some of California's first commercial fishermen netted bay shrimp, dried them on the shoreline, and then sent dried shrimp powder across the Pacific to China. At least three Chinese shrimping

villages survived into the late nineteenth century. Chinese shrimpers were the focus of thirty years of discriminatory state licensing laws, gear and seasonal restrictions, and harassment.



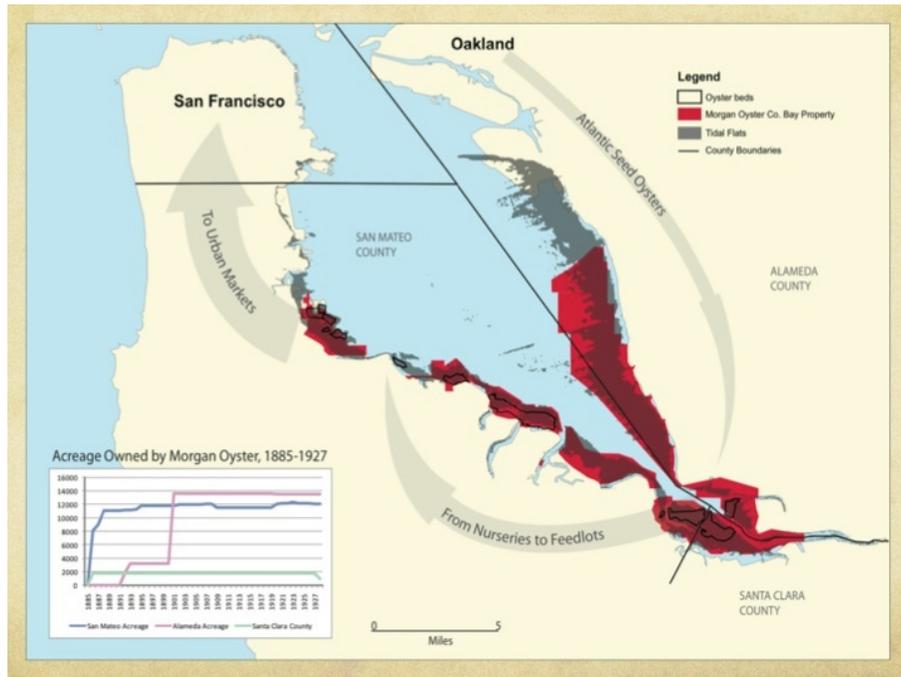
**Chinese Shrimp Fishers, San Francisco Bay, circa 1889.**  
National Archives. Courtesy National Oceanic and  
Atmospheric Administration Historical Fisheries Photograph  
Collection.

One of the California Fish Patrolmen who would persecute Chinese shrimpers was a young Jack London. Like many policemen, London joined the Fish Patrol after first leading a life of crime. At age 14 he became an oyster pirate to escape a dead-end factory job. He used the lawless spaces of the bay's edge to find freedom.

Jack London's writings, snapshots of families on the Berkeley beach, photos of Chinese shrimpers and clam diggers — these sources give tantalizing glimpses into the social world of the working majority in nineteenth century California. Most Californians then and now were city dwellers. Most, like Jack London's family, were lifelong renters who moved frequently, often to escape debt. For Jack London and for many other common people, the public spaces on the shore were essential to their economic livelihood but also to their sense of personal freedom.

Jack London described factory work as slavery, bondage he got free of only by stealing and selling oysters. London did not see it as stealing. He claimed that the oysters he pirated grew in a public space and were therefore the property of everyone. San Francisco economist, philosopher, and labor leader Henry George spoke for millions of Americans when he argued that monopoly control of land was the greatest threat to the American democracy. But George failed to see it was in the bay and not on land where monopoly was most obvious.

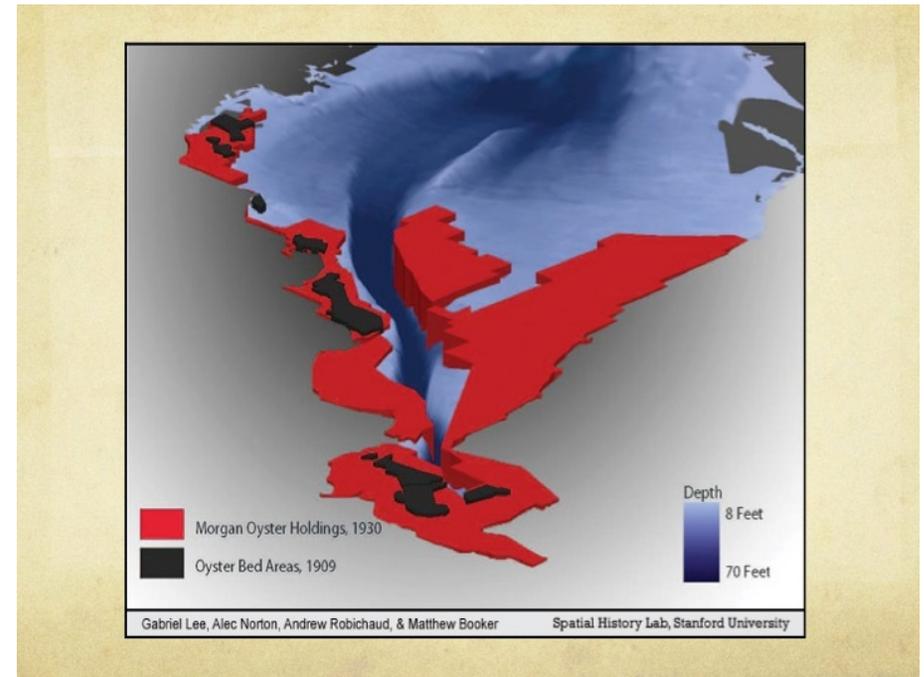
With assistance from graduate and undergraduate students at Stanford University, I mapped the process by which oyster growers first colluded to fix prices, then combined to crush competition, and finally absorbed into a single company with monopoly ownership of all the oyster beds of San Francisco Bay. We found that as the state of California sold its underwater lands to raise money for, among other purposes, the University of California, Morgan Oyster Company bought them up. All of them. Sometimes Morgan did this directly, sometimes with middlemen, and sometimes it simply claimed land that the state never sold.



Morgan Oyster Company Holdings, 1885-1930

On the eve of the Great Depression, one company not only dominated production of Atlantic oysters in California, it also controlled all the space where oysters could be grown: the definition of a monopoly.

Like many monopolies, Morgan did not last long. The company closed in 1930, selling some of its land to cement manufacturers and some to salt miners, who coincidentally were finishing their own period of consolidation into the Leslie Salt Company. Leslie ruled the bay from 1931 to the late 1970s, when they sold their remaining salt ponds to Cargill, a multinational that is one of the world's largest privately-held corporations. Like the oyster growers, Leslie fenced the public off its tens of thousands of acres.



Monopolizing Oyster Space

As Leslie drowned the marshes behind its dikes, one predictable and one very unlikely thing occurred. Predictably, the populations of salt marsh-dependent species in California plunged until a handful of animals and plants were on the verge of extinction. Unexpectedly, however, the salt ponds turned out to be a great habitat for certain species of waterfowl, including ones that people like to hunt. And so for forty years, salt mines fed and sheltered millions of waterfowl. The privatized bayshore was off-limits to hunters, fishers, and foragers, but it was a paradise for ducks.

I want to emphasize that it was an accident that salt ponds fed birds. Leslie was a corporation and it existed to maximize profits. From the beginning, it is clear that Leslie intended to make salt only until some other more profitable use could be found for its ponds. After the



Then in 2003 Senator Dianne Feinstein announced a deal to purchase most of Cargill's remaining salt ponds and to double the size of the refuge. It would be one of the largest restoration projects in history. The 2003 acquisition coincided with a powerful new push to make the shore public once again. In the 1980s and 1990s local citizens groups rallied around the idea of historical ecology — of opening up creeks that had been buried for decades, restoring marshes for endangered species, and allowing greater public access to the shoreline. Empowered with maps comparing the current bay to how it looked when Europeans arrived, local groups pushed for not just managing the refuge lands for waterfowl but rather to open the seawalls and let the tides in again. That process is ongoing, and it is far more democratic and science-based than any previous management of the bay's margin. Foragers are still largely left out of the conversation, but at least it is a conversation.

If our story ended there it would be an interesting, but hardly surprising one. Wildlife refuges elsewhere have seen similar transitions. But the future promises to exacerbate the contest over access to the bay's edge. Sea level has been rising since the 1850s, and it is projected to accelerate in coming decades, rising somewhere between 16 inches and 3 feet by 2100. This rise throws into question all that has come before. If society does nothing the bay appears likely to refill to its 1849 shoreline, flooding the ports, factories, and freeways built since the gold rush.

But this society is not going to just let nature take its course. Americans are too invested in the current users of the shore, and recently they have doubled down. Both Google and Facebook, the stars of Silicon Valley, have recently built immensely expensive headquarters on filled land in the path of rising tides. Both expect the public to defend their property at public expense.



This image is based on a map produced by the San Francisco Bay Conservation and Development Commission

Local environmental organizations use historical ecology to make a claim for a previous system that worked. Implicitly they also claim a voice for the everyday people of this place in managing the bayshore. But do environmentalists remember the inhabited bayshore, the human uses and purposes of the nineteenth century bay?

What has been eroded from our memories of the shoreline is that human bay, the many uses that an open access bayshore offered. Those older uses are layers of accreted history — marshes, oyster beds, salt mines, and wildlife refuge. Until it is an edible landscape once again, this place cannot be considered whole, or healthy.

If we cannot even visualize what that past looked like, it becomes much more difficult to imagine its future.

Originally published by Matthew Booker on [June 11, 2012](#). Revised for the *Journal of Digital Humanities* September 2012.

Copyright for the images above by the Spatial History Project at Stanford University.

MELISSA TERRAS

# The Impact of Social Media on the Dissemination of Research: Results of an Experiment

In September 2011 I returned to work after a year on maternity leave. Many things needed sorting out, not least my digital presence at my home institution, which had switched to a content management system that seamlessly linked to University College London's open-access repository, "[Discovery](#)." The idea was we should upload open-access versions of all our previously published research, and link to it from our home pages, to aid in dissemination.

There is no doubt that this type of administrative task is tedious. To break up the monotony of digging out the last previous version prior to publication of my 26 journal papers (we put up a last-but-one copy to get around copyright issues with journals) I decided to blog the process. I wrote a post about each paper, or each research project that had spawned papers. I wanted to tell the stories behind the research – the things that don't get into the published versions. I also set about methodically tweeting about these research papers, as they went live, going through my back catalogue in reverse chronological order.

What became clear to me very quickly was the correlation between talking about my research online and the spike in downloads of my

papers from our institutional repository. A game that had spurred me to carry out an administrative task was actually disseminating my research quite effectively. So this, in turn, became the focus of the blog posts that are featured here.

The first, "What Happens When You Tweet an Open-Access Paper" discusses the correlation between talking about an individual paper online, and seeing its downloads increase. The second, "Is Blogging and Tweeting About Research Papers Worth It? The Verdict" discusses the overall effect of this process on all my papers, highlighting what I think the benefits of open access are. In the final post, "When Was the Last Time You Asked How Your Published Research Was Doing?" I talk about the link between publishers and open access, and how little we know about how often our research is accessed once it is published.

More than 20,000 people have now read these three online posts. It is evident to me that academics need to work on their digital presence to aid in the dissemination of their research, to both their subject peers and the wider community. These blog posts provide the evidence to prove this.

## What Happens When You Tweet an Open-Access Paper

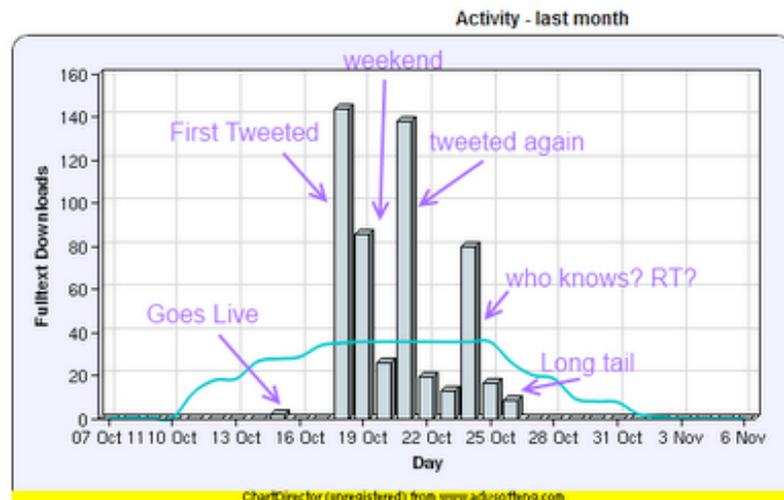


Chart of the Download Activity for "Digital Curiosities" from UCL "Discovery."

So a few weeks ago, I tweeted and [posted](#) about this paper:

Melissa Terras, "Digital Curiosities: Resource Creation Via Amateur Digitisation," *Literary and Linguistic Computing*, 25.4 (2009): 425 - 438. Available in [PDF](#).

I thought it worth revisiting the results of this. Is it worth me digging out the full text, [running the gamut with the UCL repository](#), and trying to spend the time putting my previous research online? Is open access a gamble that pays, and if so, in what way?

Prior to me blogging and tweeting about the paper, it was downloaded twice (not by me). The day I tweeted and blogged it, it immediately got 140 downloads. This was on a Friday; on the Saturday and Sunday it

got downloaded, but by fewer people. On Monday it was retweeted and the paper received a further 140 or so downloads. I have no idea what happened on the 24th of October – someone must have linked to it? Posted it on a blog? Then there were a further 80 downloads. Then the traditional long tail, then it all goes quiet.

All in all, it's been downloaded 535 times since it went live, from all over the world: USA (163), UK (107), Germany (14), Australia (10), Canada (10), and the long tail of beyond: Belgium, France, Ireland, Netherlands, Japan, Spain, Greece, Italy, South Africa, Mexico, Switzerland, Finland, Denmark, Norway, Sweden, Portugal, Europe, United Arab Emirates, "unknown".

Worth it, then? Well there are a few things to say about this.

- I have no idea how many times it is read, accessed, or downloaded in the journal itself. So seeing this – 500 reads in a week! – makes me think, "wow, people are reading something I have written!"
- It must be all relative, surely. Is 500 full downloads good? Who can tell? All I can say is that it puts it into the top ten – maybe top five – papers downloaded from the UCL repository last month (I won't know until someone updates the webpage with last month's statistics).
- If I tell you that the most accessed item from our department ever in the UCL repository, which was put in there five years ago, has had 1,000 full text downloads, then 500 downloads in a week isn't shabby. They didn't blog or tweet it, it's just sitting there.
- There is a close correlation between when I tweet the paper and downloads.

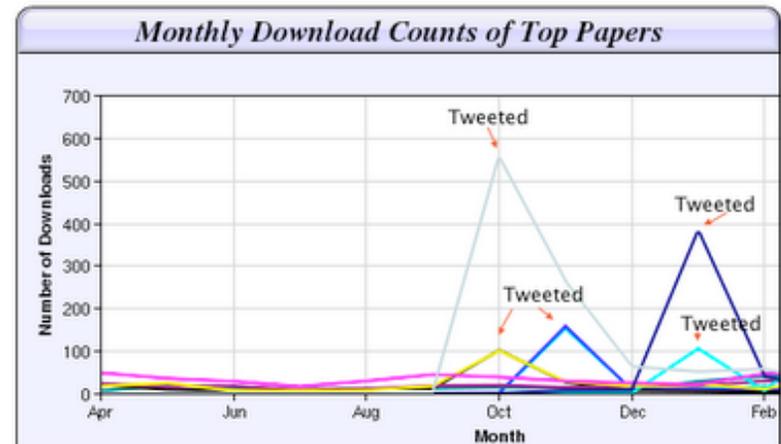
- There can be a compulsion to start to pay attention to statistics. Man, it gets addictive. But is this where we want to be headed: academia as X-factor?

Ergo, if you want people to read your papers, make them open access, and let the community know (via blogs, twitter, etc.) where to get them. Not rocket science. But worth spending time doing. Just don't develop a stats habit.

\* \* \*

The updated UCL [statistics page for downloads](#) shows that "Digital Curiosities" was the fifth most downloaded paper in the UCL repository in October 2011. Yeah, I'm up there with fat tax, seaworthiness, preventative nutrition, and the peri-urban(?) interface. The Digital Curation Manager at UCL, Martin Moyle, has been in touch to confirm that 6,486 of the 224,575 papers in the repository have downloadable full text attached.

## Is Blogging and Tweeting About Research Papers Worth It? The Verdict



- Digital Curiosities: Resource Creation Via Amateur Digitisation
- Enabled backchannel: conference Twitter use by digital humanists Not Me
- Framework for effective public digital records management in Uganda
- Library and information resources and users of digital resources in the huma
- A Virtual Tomb for Kelvingrove: Virtual Reality, Archaeology and Education
- What do faculty and students really think about e-books? Not me
- Documentation and the users of digital resources in the humanities
- Classification in British public libraries: a historical perspective Not me
- Teaching TEI: The Need for TEI by Example
- Should we just send a copy? Digitisation, Use and Usefulness

**Guess When I Tweeted My Papers? Top Ten Downloaded Papers From My Department in the Last Year, Seven of Which Include Me in the Author List**

In October 2011, I [began](#) a project to make all of my 26 articles published in refereed journals available via UCL's Open Access Repository, [Discovery](#). I decided that as well as putting them in the institutional repository, I would write a blog post about each research

project, and tweet a link to download the paper. Would this affect how much my research was read, known, discussed, distributed?

I wrote about the stories behind the research papers – from becoming [so immersed in developing 3D that you start walking into things in real life](#), to [nearly barfing over the front row of an audience's shoes](#) whilst giving a keynote, to [passive aggressive notes from an archaeological dig](#) that take on a digital life of their own. I gave a run down, in roughly reverse chronological order, of the twelve or so projects I've been involved in over the past decade that resulted in published journal papers. Along the way, I wrote a little bit about [the difficulties of getting stuff into the institutional repository](#) in the first place, but the thing that really flew was my post on [what happens when you blog and tweet a journal paper](#), showing (proving?) the link between blogging and tweeting and the fact that people will download your research if you tell them about it.

So what are my conclusions about this whole experiment?

Some rough stats, first of all. Most of my papers, before I blogged and tweeted them, had one to two downloads, even if they had been in the repository for months (or years, in some cases). Upon blogging and tweeting, within 24 hours, there were on average 70 downloads of my papers. Seventy. Now, this might not be internet meme status, but that's a huge leap in interest. Most of the downloads followed the trajectory I described with the downloads of "[Digital Curiosities](#)," in that there would be a peak of interest, then a long tail after. I believe that the first spike of interest from people clicking the link that flies by them on twitter (which was sometimes retweeted) is then replaced by a gradual trickle of visitors from postings on other blogs, and the fact that the very blog posts about the papers make them more findable when the subject is googled. People read the blog posts – I have about 2,000 visitors to my blog a month, 70% new, with an average time on

the site of one minute and five seconds. People come here, tend to read what I have written, and seem to be clicking and downloading my research papers.

The image above shows the top ten papers downloaded from my entire department over the last year. There were a total of 6,172 downloads from our department ([UCL Department of Information Studies](#) is one of the leading iSchools in the UK). Look at the spikes. That's where I blog and tweet about my research. I'm not the only person producing research in my department (I think there are eighteen current staff members and a further twenty or so who have moved on but still have items in the institutional repository) but I'm the only person who has gone the whole hog on promoting their research like this. You will see that seven out of ten of the most downloaded papers from my Department in the last calendar year have me in the author list. As a clue, I don't know anything about Uganda, e-books, or classification in public libraries. In the last calendar year 27 out of the top 50 downloads in our department feature me (as a rough guide, I get about 1/3 of the entire downloads for my department). My stuff isn't better than my colleagues' work. They're all doing wonderful things! But I'm just the only one actively promoting access to my research papers. If you tell people about your research, they look at it. Your research will get looked at more than papers which are not promoted via social media.

Some obvious points and conclusions. Don't tweet things at midnight, you'll get half the click throughs you get during the day when people are online. Don't tweet important things on a Friday, especially not late – people do take weekends and you can see a clear drop off in downloads when the weekend rolls around and your paper falls a bit flat, as you sent it on its way on social media at the wrong time. The best time is between 11am and 5pm GMT, Monday to Thursday in a

working week. I have the stats here somewhere to prove it. I won't write it up, though, as it's pretty predictable. It is important to note that just putting links on twitter isn't enough, you have to time it right. The [Discovery twitter account](#) regularly posts an automated list of the really interesting things people have been looking at ... at 10pm on a Friday night. I only know as I'm regularly sad enough to still be on twitter at that time, but I suspect if they tweeted the papers through the day during the working week ... well, you guess what would happen.

The paper that really flew – "Digital Curiosities" – has now been downloaded over a thousand times in the past year. It was the [16th most downloaded paper from our entire institutional repository in the final quarter of 2011](#), and the third most downloaded paper in UCL's entire Arts Faculty in the past year. It's all relative really – what does this really mean? Well, I can tell you that this paper was the most downloaded paper in 2011 in *Literary and Linguistic Computing (LLC)* Journal, where it was published (and where [it lives behind a paywall](#) apart from being available free from Discovery). *LLC* is the most prestigious journal in the discipline I operate in, Digital Humanities. The entire download count for this paper from *LLC* itself, which made it top paper last year? 376 full text downloads. There have been almost three times that number of downloads from our institutional repository. What does this mean? What can we extrapolate from this? I think it's fair to say that it's a really good thing to make your work open access. More people will read it than if it is behind a paywall. Even if it is the most downloaded paper from a journal in your field, open access makes it even more accessed.

However, I might just have written a nice paper that caught peoples' interest; there are, after all, no controls to this, are there? No controls! How can we tell if papers would fly without this type of exposure? Well.

Erm. I might have not tweeted one or two papers to see the difference between tweeting and blogging about papers and not doing so. Take the LAIRAH (Log Analysis of Internet Resources in the Arts and Humanities) project, which I [wrote about here](#). We actually published four papers from this research. I tweeted and promoted three of them actively. One I didn't mention to you. Here are the download counts. Guess which one I didn't circulate?

- "[Library and information resources and users of digital resources in the humanities](#)": 297 downloads.
- "[Documentation and the users of digital resources in the humanities](#)": 209 downloads.
- "[If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data](#)": 142 downloads.
- "[The Master Builders: LAIRAH Research on Good Practice in the Construction of Digital Humanities Projects](#)": 12 downloads.

The papers that were tweeted and blogged had at least more than 11 times the number of downloads than their sibling paper which was left to its own devices in the institutional repository. *Q.E.D.*, my friends. *Q.E.D.*

I can't know if the downloaded papers are read though, can I? The only way to do so is to enter the murky world of citation analysis. The trouble with this is the proof of the pudding will come to light in a few years time – if someone reads something of mine now and decides to cite it, it's going to take one or even two years (or more) for it to appear in my citation list. So, I'll be keeping an eye on things, not too seriously as we all know things like H-index are problematic. Just for the record, at time of writing, I have [218 citations, according to Google Scholar](#).

My H-index is 8, and my i10 index is 5, which is ok for a relatively young Humanities scholar (I'm still technically an Early Career Researcher for another year, as defined by the UK funding councils). "Digital Curiosities" only has three published citations to date. Three published citations. Remember, it's been downloaded over 1,300 times, between *LLC* and our repository. Will this citation count grow? Will I be able to demonstrate, over the next few years, that retweeting leads to citation? Will I be able to tell how people came across my research? We'll see. Don't worry, I'll blog it if I have anything to say on this.

I also know nothing about how many times my other papers are downloaded from the websites of published journals, or consulted in print in the library. The latter, no one can really say much about – but the former? It seems strange to me that we write articles (without being paid) and we get them published by people who make a profit on them, yet we don't even know – usually – how many downloads they are getting from the journals themselves. The only reason I know about the *LLC* statistics is because I am good friends with the [editor](#). So, there are obvious advantages to being able to monitor my own downloads from my institutional repository. It's been a surprise to me to see what papers of mine are of interest to others. (Should that drive my research direction, though?)

The final point to make is that people don't just follow me or read my blog to download my research papers. This has only been part of what I do online – I have more than 2,000 followers on twitter now and it has taken me over three years of regular engagement – hanging out and chatting, pointing to interesting stuff, repointing to interesting stuff, asking questions, answering questions, getting stropky, sending supportive comments, etc. – to build up an "audience" (I'd actually call a lot of you friends!). If all I was doing was pumping out links to my published stuff, would you still be reading this? Would you have read

this? Would you keep reading? My blog is similar: sure, I've talked about my research, but I also post a variety of other content, some silly, some serious, as part of my academic work. I suspect this little experiment only worked as I already had a "digital presence," whatever that may mean. All the numbers, the statistics. Those clicks were made by real people.

So that would be my conclusion, really. If you want people to find and read your research, build up a digital presence in your discipline, and use it to promote your work when you have something interesting to share. It's pretty darn obvious, really:

If (social media interaction is often) then (open access + social media = increased downloads).

What next? From now on, I will definitely post anything I publish straight into our institutional repository, and blog and tweet it straight away. After all, the time it takes to undertake research, and write research papers, and see them through to publication is large; the time it takes to blog or tweet about them is negligible. This has been a retrospective journey for me, through my past research, at a time when I came back from a period of [leave](#). It's been fun to get my act together like this – in general I needed to sort out my online systems at UCL, so it gave me some impetus to do so. But it has shown me that making your research available puts it out there – and as soon as I have something new to show you, you'll be the first to know.

## When Was the Last Time You Asked How Your Published Research Was Doing?

Whilst writing up my thoughts about [whether blogging and tweeting about academic research papers was "worth it,"](#) the one thing that I found really problematic was the following:

I also know nothing about how many times my other papers are downloaded from the websites of published journals, or consulted in print in the Library. The latter, no one can really say much about – but the former? It seems strange to me that we write articles (without being paid) and we get them published by people who make a profit on them, yet we don't even know – usually – how many downloads they are getting from the journals themselves.

That's true enough, I thought. But whose fault is it that I don't know about access statistics for journals I have published in? Have I ever *asked* for the access statistics for how many times my papers have been downloaded from the journals they are published in? Has anyone?

So, Reader, I asked for some facts and figures regarding the circulation of journals and the download statistics of my papers.

I have to say that the journals were really very helpful, and forthcoming, if surprised. "I imagine the publishers would be happy to tell an author the cumulative downloads for their papers ... So far as I know, you are the first author ever to ask ... certainly the first to ask me," said [David Bawden](#), editor of the [Journal of Documentation](#) (*JDoc*). Jonas Söderholm, editor of [HumanIT](#), highlighted some of the issues journals will face if people start asking this kind of question, saying:

A reasonable request and we would gladly assist you. Unfortunately we do not have direct access to server logs as our web site is hosted as part of the larger University of Borås web. We will take your request as a good

excuse to check into the matter though, and also review our general policy on log data.

Most journals got back to me by return of email, telling me immediately what they knew and were very aware of the limitations of their reporting mechanisms, for example whether or not the figures excluded robot activity, the fact that how long the user stays on the website is not known so accidental click-throughs are undetermined, etc. Such caveats were explained in detail. Emerald, the publishers of *JDoc* and [Aslib Proceedings](#), were not comfortable with giving me access to wider statistics about their general readership numbers, given this could be commercially sensitive information, which is understandable; they were very happy to give me the statistics relating to my own papers, though.

The only journal not to get back to me was [LLC](#), published by Oxford University Press (the editor replied to say he was not sure he had access to these statistics, but would ask). This is ironic, given I'm on the editorial board. I'll press further, and take it to our summer steering-group meeting.

I suspect that the actual statistics involved are only really very interesting to myself. I had originally planned to make comparisons with the amount of downloads from [UCL Discovery](#) (open access is better, folks! etc.), but I think the picture is foggier than that. What this exercise does do is highlight the type of information that, as authors, we don't normally hear about, which can be actually quite interesting for us, as well as stressing the complex relationship between open access and paywalled publications. Here are some details:

- One of my papers published in *JDoc*, "[Enabled backchannel: conference Twitter use by digital humanists](#),"<sup>[1]</sup> was downloaded 804 times from the *JDoc* website during 2011, and was number 16

in the download popularity list that year. The total number of paper downloads from *JDoc* as a whole during that year was 123,228. Isn't that interesting to know? I have a top twenty paper in a really good journal in my discipline! Who knew? It has now been downloaded 1,114 times from their website. In comparison, there have been 531 total downloads of that paper from UCL Discovery in the past six months. But the time frame for comparison of downloads with the open-access copy from Discovery isn't the same, so comparing is problematic – and there are *more* downloads from the subscription journal than from our open-access repository. Still, it shows a healthy amount of downloads, so I'm happy with that.

- The [\*Art Libraries Journal\*](#), only available in print, not online, were quick to tell me that the journal is distributed to 550 members: 200 going abroad to Libraries/Institutions, 150 sent to UK Personal members, and 200 going to UK Libraries/Institutions. My paper published there, "[Should we just send a copy? Digitisation, Use and Usefulness](#),"[2] has had 205 downloads in the last six months from UCL Discovery, so I perceive that as a really good additional advert for open access: the print circulation is fairly limited, but the open-access copy is available to all who want it.
- My paper in the open-access [\*International Journal of Digital Curation\*](#), "[Grand Theft Archive: a quantitative analysis of the current state of computer game preservation](#),"[3] was downloaded 903 times in 2009 out of the 53,261 times the full text of a paper was accessed. (The average was 476, with standard deviation 307). In 2010 the paper accounted for 919 out of the 120,126 times the full text of a paper was accessed. (The average was 938, with standard deviation 1,045.) That compares to only 85 downloads from the UCL repository, but hey, it's freely available online *anyway*, without having to revert to an open-access copy in

an institutional repository. It might be worth drawing from this that copies of papers in institutional archives are only really used when the paper isn't available anywhere else, but you would hope that would be obvious, no?

- [\*Internet Archaeology\*](#) has an online page with their [download statistics readily available](#) (how I wish all journals would do this). The journal gets around 6,200 page requests per day. But since article size varies widely, with some split into hundreds of separate HTML pages, it is difficult to know how meaningful this is. I was sent a spreadsheet of the statistics from my paper published there, "[A Virtual Tomb for Kelvingrove: Virtual Reality, Archaeology and Education](#),"[4] which suggests that there have been 2,083 downloads of the PDF version of the paper from behind the paywall since 2001 (but some may be missing due to the way the reporting mechanism is set up) with none in the past year (compared to 276 downloads of this from UCL Discovery in the past six months, so many more from our institutional repository comparing like periods). The HTML version of the table of contents has been consulted 16,282 times since 2001 (this is freely available to all comers) but there have been 67,525 views of all files in the directory since then – but since the paper is comprised of hundreds of individual files, it is difficult to ascertain readership. Judith Winters, the editor of *Internet Archaeology*, notes "It is curious that when the journal went open access for about 2 weeks towards the end of last year, the counts did increase but not dramatically so" – so when a non-open-access journal throws open its doors for a limited time (*Internet Archaeology* did this to mark [Open Access Week](#) last year) it is not like access figures go wild. That's really interesting, in itself.

It has been fascinating, for me, to see the (mostly positive) reactions publishers have to being approached about this – and surprising that not more people have actually asked publishers about these statistics. We are giving away our scholarship to publishers, in most cases; shouldn't we get to know how it fares in the wide, wide world? As citation counts, and H-indexes, and "impact" become increasingly important to external funding councils and internal promotion procedures within universities, why would journal publishers not make this information available to authors?

Will you need this type of information for the next grant proposal, or internal promotion, you chase? Why would you not be interested in how your research flies? But journal publishers will only start providing authors with this kind of information routinely if enough scholars start to ask about it, and it becomes part of the mechanics of publishing research – particularly when publishing research online.

So if you have published in a print journal which has an online presence, or in an online journal, drop them an email to ask politely how your downloads are going. Do it. Do it now. Ask them. Ask them!

Originally published by Melissa Terras on [November 7, 2011](#), [April 3, 2012](#), and [May 16, 2012](#). Revised with new introduction for *Journal of Digital Humanities* September 2012.

**Notes:**

[1] Claire Ross, et al. "[Enabled Backchannel: Conference Twitter Use by Digital Humanists](#)," *Journal of Documentation* 67.2 (2011): 214 - 237.

[2] Melissa Terras, "[Should we just send a copy? Digitisation, Use and Usefulness](#)," *Art Libraries Journal* 35.1 (2010).

[3] Paul Gooding and Melissa Terras, "[Grand Theft Archive: a quantitative analysis of the current state of computer game preservation](#)," *International Journal of Digital Curation* 3.2 (2008).

[4] Melissa Terras, "[A Virtual Tomb for Kelvingrove: Virtual Reality, Archaeology and Education](#)," *Internet Archaeology* 7 (1999).

# ORBIS

## PROJECT OVERVIEW

### **ORBIS: An Interactive Scholarly Work on the Roman World**

Elijah Meeks & Karl Grossner . . . . . 40

## CASE STUDY

### **Modeling Networks and Scholarship with ORBIS**

Elijah Meeks & Karl Grossner . . . . . 42

## REVIEW

### **Review of ORBIS**

Stuart Dunn . . . . . 48

# ORBIS: An Interactive Scholarly Work on the Roman World



There are, broadly speaking, four expressions of digital humanities scholarship: research utilizing digital objects, tools, and methods; the publication of the products of that research; the creation, extension, and annotation of digital archives; and the development or improvement of digital tools. The boundaries between these are fluid;

in particular, some products of digital humanities scholarship blur the line between archive, tool, and publication. [ORBIS – the Stanford Geospatial Network Model of the Roman World](#),<sup>[1]</sup> is one such hybrid. It is an interactive scholarly work (ISW) that comprises a digital archive of sites and routes, a tool for exploring Roman transportation, and an argument about the dynamic shape of the Roman world and the nature of transport within it. Furthermore, it makes a methodological argument that its representations – its computational model and visualizations – are a useful means for reasoning about those aspects of Roman history.



One of the forms of expression of ORBIS is a route-finding interface that affords a user the opportunity to examine the distance, duration, and expense of a trip between any two sites represented within the system.

The actual form of ORBIS also blurs some of these boundaries. It is a website with both static and interactive components. Its textual content itself forms an academic article, such as might appear in a print journal. An application programming interface (API) in development will allow users to access the data directly and

programmatically, and the site's interactive components, [discussed in more detail here](#), provide several perspectives on the model itself.

ORBIS crosses disciplinary boundaries as well. It integrates several geographic network analysis methodologies like least cost paths, cost surfaces, and distance cartograms into a historical project, and it is a platform for research using traditional network analysis in the context of world systems theory.

Built over a nine month period, ORBIS combines location data drawn primarily from [the Pleiades project](#),<sup>[2]</sup> route information from the *Barrington Atlas*,<sup>[3]</sup> and established courses of historically navigable rivers. Together these represent the terrestrial transportation network well, and they were joined with a novel maritime transportation network modeled on the performance of Roman-era ships and using modern wind and sea data. It includes a temporal component that is both detailed and highly abstract. In the first case, ORBIS is extremely time-aware in that both terrestrial and sea route availability and speed can change based on the month of travel. However in another sense ORBIS represents a purely timeless transportation network of Rome, given that it contains no annotation of route availability based on the period of the regime. Taken altogether, it gives prospective readers an expression of the dynamic and contingent shape of Roman world.

#### Notes:

[1] <http://orbis.stanford.edu>

[2] <http://pleiades.stoa.org/>

[3] When available. Certain routes, such as the desert routes, were missing, and the entire panel for Greece shows no roads, almost as if to argue that Greece was an eco-utopia not ravaged by cruel roadmakers. Available online at [http://www.unc.edu/depts/cl\\_atlas/](http://www.unc.edu/depts/cl_atlas/). Also available in print: Richard J.A. Talbert (ed.), *Barrington Atlas of the Greek and Roman World* (Princeton: Princeton University Press, 2000).

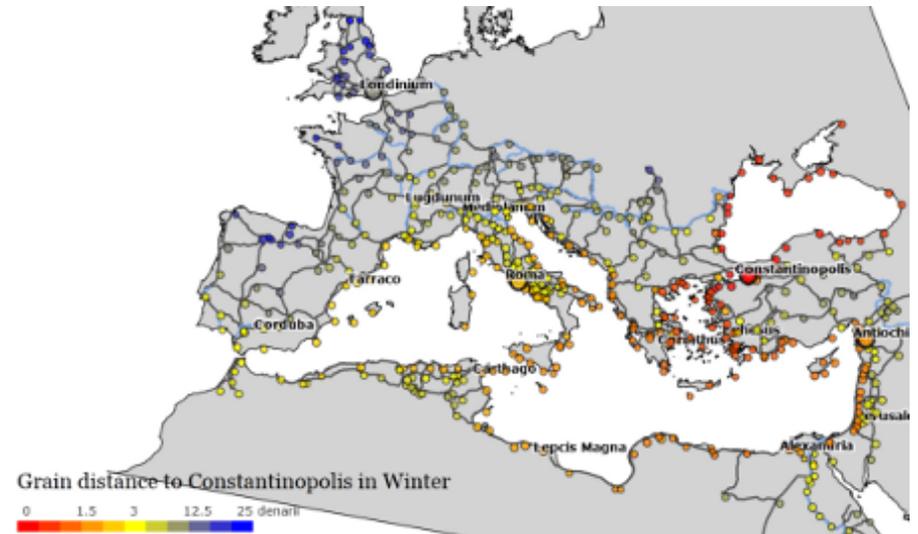
# Modeling Networks and Scholarship with ORBIS

ORBIS is self-defined as a historical transportation network model of the Roman world, but while it is fundamentally a computational model, it is also part of a genre of objects that could be referred to as *interactive scholarly works*, and so in explaining how to use ORBIS it becomes necessary to also engage with interactive scholarly works in general. What follows is not solely an explanation of ORBIS, but an exploration of the genre in a broader sense, providing both pragmatic and theoretical examples of the usage of both.

ORBIS offers three primary perspectives on the transport of goods and movement of individuals in the Roman Empire: global, strategic, and situated. Each will be more or less suitable depending on the scale of inquiry. These views into the underlying model are presented by means of (respectively) [a dynamic distance cartogram](#), [a route mapping feature](#), and [a first-person way-finding application](#).

The distance cartogram demonstrates the dynamics of the general shape and size of the Roman world based on season and travel priority (choice of fastest or cheapest route). Geographic distances from Constantinople, Rome, London, and Antioch to the rest of the empire

are transformed to time and expense distances for either January or June. Although computational cost permits offering only four sites at this time, each of these transfigurations imparts significant information about the remaining sites in relation to the four and each other.

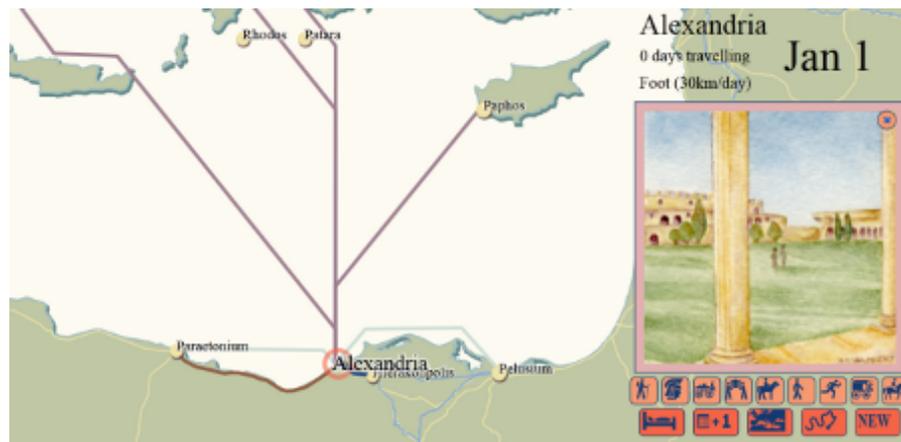


The ORBIS dynamic distance cartogram affords users the capability of seeing the entire Roman world system at once, distorted based on time of year and priority of travel.

At the next scale of inquiry, the ORBIS route mapper allows users to examine the costs and specific geometries of travel between any two of 650 sites based on a host of options. The use of the route map is explained in detail in the “Using ORBIS” section of the site, which includes a video tutorial. The results of this route-finding exploration can be downloaded in two standard file formats and integrated into other research (as KML for mapping or CSV for simple graphing or statistical analysis).



One of the forms of expression of ORBIS is a route-finding interface that affords users the opportunity to examine the distance, duration, and expense of a trip between any two sites represented within the system.



A view of the ORBIS model from the perspective of the site of Alexandria, showing the variety and nature of connections between Alexandria and other sites in the model.

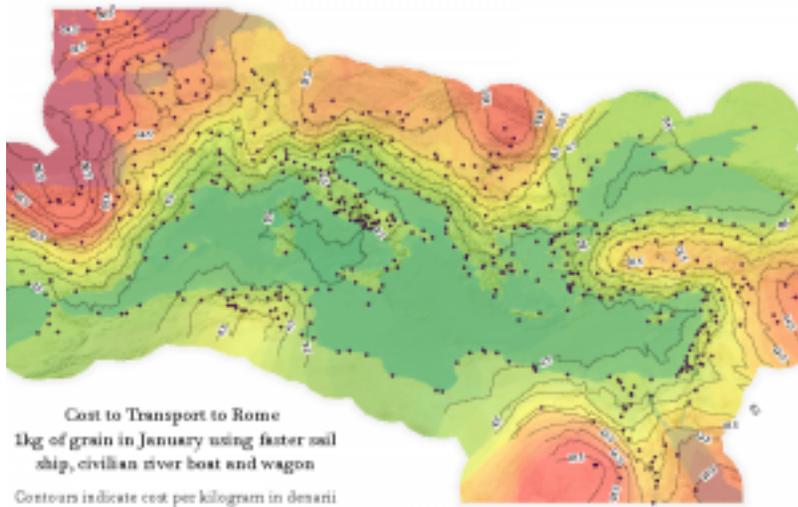
A third, local view of the network is provided with the embedded ORBIS|via application, which allows users to navigate around the Roman network from site to site. As above, travel times are based upon

the month and the mode or vehicle selected. The implementation of this “situated view” is to date fairly rudimentary, built as a proof-of-concept investigation of various aesthetic and functional forms, including those found in early video games.



The text and static images of much of the ORBIS site are available to download, formatted as a traditional journal article. While this may seem anachronistic, providing traditional representations of these kinds of works improves the likelihood of reuse in future scholarship.

There is a fourth mode of presentation, which could be easily overlooked given its traditional nature: virtually all the text and figures of ORBIS have been assembled verbatim as an article, downloadable from the site in PDF format. Arguably, scholarly digital products like the downloadable KML spatial data describing routes in ORBIS can be embedded in others’ work in much the same way that written arguments are with more traditional citation methods.



An isophoretic map of the Roman world, which uses cost contours to show the price to ship goods to Rome from all sites modeled in ORBIS. An isophoretic map requires not only the computation of numerous least-cost-paths, but the interpolation of such costs into a surface followed by the creation of contours based on that surface. This would require too much processing power to be made real-time and web-accessible on the ORBIS site.

Integral to this mode and capable of being engaged on their own are the series of static maps and other visualizations that populate the site. While dynamic and interactive elements provide significant capacity for users to engage with this type of content, static visual representations of data still provide the best method of communication in many cases, and are the only method to represent data that is too computationally expensive to provide in a web-accessible format. Many of these maps are available in the **Map Gallery** tab of **Mapping ORBIS**. Among these are not only maps of the Roman world based on the ORBIS model, but also maps and data visualization indicating how the ORBIS model functions.

## Engagement with Modeled Arguments

Although ORBIS can be viewed as a tool for understanding travel in the Roman world, as a model it is also fundamentally a claim about the dynamic shape of the Roman world and the nature of transport within it. As Willard McCarty notes, “a model is by nature a simplified and therefore fictional or idealized representation, often taking quite a rough-and-ready form.”<sup>[1]</sup> This representation is an argument or proposition that can be amended, extended, challenged, or refuted in the same manner of scholarly discourse involving monographs and journal articles and as noted, by integration of digital products like spatial and tabular data in other works. More ambitious scholars can integrate the code for ORBIS, in whole or in part, into their own systems for pre-modern transportation networks around the world. However, given the formal and explicit nature of a computational model, scholars can potentially pursue a more sophisticated engagement with individual components of it without focusing on either extreme.

McCarty writes of modeling being the means to achieve, “a computing that is *of* as well as *in* the humanities: a continual process of coming to know by manipulating representations.” Such direct integration and interaction with the model in unexpected ways requires digital interfaces to it, and is part of a movement toward interoperability between geospatial models reflecting a growing trend in GIS.<sup>[2]</sup> One of these interfaces, the API, allows programmatic consumption of data and results, permitting scholars to not only embed specific elements of one model or dataset into their own, but also to cite these elements in their scholarship. APIs still require a level of programming knowledge that may mitigate engagement, and so as an extension of that concept, and possibly more useful to many digital humanities scholars, the next phase of ORBIS provides both an API<sup>[3]</sup> and an interactive “exploded”

representation of the model. This will include the means for users to enter their own values for parameters such as vehicle speed, seasonal weighting, and so on, without requiring the level of programming knowledge necessary to use an API.

## Interactive Scholarly Works in Education

Like other secondary sources, interactive scholarly works (ISWs) can be formatted or presented in such a way as to serve as textbooks; that is, teaching resources around which entire courses or sets of course units can be structured. In the case of ORBIS, development of such digital resources followed naturally from a research project; others might be designed primarily for educational purposes from their inception. ISWs such as ORBIS are particularly well-suited to the emerging “massively on-line course” (MOOC) setting. As it stands, much of the course material for MOOCs is dynamic but not interactive – most often it is interactive only in features like polls and forums.

When integrated with lesson plans and exercises designed to stimulate discussion and active engagement – using simulations, for example – an ISW could provide the means to transition MOOCs from a relatively simple ‘lecture-plus-homework’ framework to something more nearly like seminars. We don’t believe computing resources are an impediment to adding functionality; when ORBIS received extremely high traffic during particular periods in its first month, performance was more than adequate with only modest server resources. While ORBIS was not designed as a MOOC, it could easily be extended to allow scenarios and content to change over the course of a class, and even the creation of content on the part of users, fulfilling the need for essays and testing, could be integrated into the site architecture.

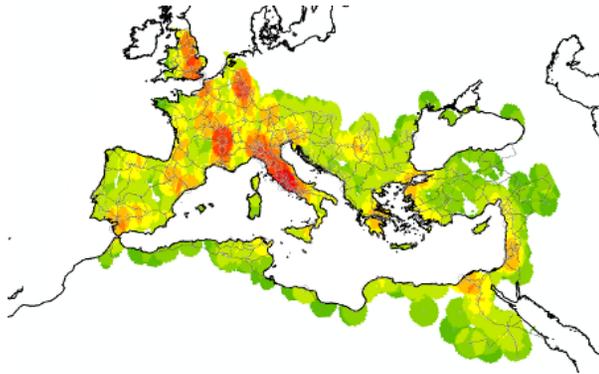
1.	calculate route	30.72%
2.	Mapping-Map	17.70%
3.	Mapping-Top	17.23%
4.	Mapping-Cartogram	5.90%
5.	Using	3.89%
6.	Using-Overview	3.50%
7.	Introducing	2.31%
8.	Mapping-Gallery	2.21%
9.	Building	1.56%
10.	Understanding	1.55%
11.	Applying	1.53%
12.	Understanding-Intro	1.47%
13.	Building-Historical	1.44%
14.	Grain distance to Rome in Summer	0.98%
15.	Travel time to Rome in Winter	0.90%
16.	Travel time to Rome in Summer	0.85%
17.	Grain distance to Rome in Winter	0.83%
18.	Credits	0.45%
19.	Using-Examples	0.40%
20.	road	0.40%

Overall site usage of ORBIS based on the activity of users. Note that this can represent both traditional page-view style metrics as well as engagement with particular functions of the site.

## Web Analytics and Pedagogy

ISWs lend themselves to tracking student usage at both coarse and fine resolution. We argue this can help teachers of these courses to better understand how their students are engaging with the material. At a coarse resolution, simple web analytics can demonstrate the amount of time spent on a site or on any of its pages. At a finer resolution, individual actions and activity paths can be recorded. As ORBIS was not designed to track this kind of usage from its inception, we can only provide a few examples based on the rudimentary analytics currently available on the site.

Simple analytics can demonstrate the amount of time spent on a website, as well as the actions performed there. These “events” can consist not only of visiting a particular tab but of performing a specific activity, such as calculating a route in the route finder, traveling by road in ORBIS|via, or distorting the cartogram according to one of the various options available.



**A heat map demonstrating the usage of ORBIS|via based on site selection. Such heat maps or other representations of usage can easily be correlated with proficiency demonstrated in more traditional areas, such as exams and papers.**

## Engaging the general public

The extraordinary traffic to the ORBIS site<sup>[4]</sup> indicates to us that people really like ISWs. Although the potential popularity will vary considerably depending on topic and the range of functions, affordances provided in them mean that scholars are able to simultaneously release a rigorous scholarly work and a popular scholarly work. Built for one audience, the underlying model can address many, and requires little effort to transition between them.

## Conclusion

A well-designed interactive scholarly work is amenable to multiple points of view as well as multiple points of interaction, blurring the line between academic and popular scholarship, as well as research, archive, and pedagogical resource. While ORBIS provides its creators with much of that capacity, it does not yet provide it to everyday users. It is really in this final sense that ORBIS, while a tool, is very much a piece of scholarship.

Future extension of ORBIS will provide more opportunities for scholars and other audiences to use it in these and as-yet-unforeseen manners, and making the inner working of models accessible is an important and challenging design issue for digital humanities scholars. Ultimately, models are best understood by those that create and engage with them, and so ORBIS is at its best as a work when seen not just as *orbis.stanford.edu* but also as a set of functions, code, data, and methodologies to be replicated and extended elsewhere. To treat such objects only as tools, and implicitly capable only of providing that which they were designed to provide, undercuts the possibilities of advancing the use of models and modeling in the humanities. This is why it is proper to treat interactive scholarly works not only as

computational models but also as publishing models, with multiple forms of expression creating a distinct genre.

**Notes:**

[1] Willard McCarty, "Modeling: A Study in Words and Meanings" in Susan Schreibman, Ray Siemens, and John Unsworth, eds., *A Companion to Digital Humanities* (Oxford: Blackwell, 2004). <http://www.digitalhumanities.org/companion/>.

[2] G. Geller and W. Turner, "The Model Web: A Concept for Ecological Forecasting," in *Proceedings of the Geoscience and Remote Sensing Symposium*, 2007. IGARSS 2007. IEEE International: 2469–72.

[3] <http://orbis.stanford.edu/api/>

[4] Over 400,000 visitors to a scholarly research publication in its first three months.

# Review of Orbis

Quantitative methods matter in the digital humanities, and are coming to play an ever more prominent role in their discourse. This is driven, in part, by a great increase in interest in simulation modelling. Often borrowing methods and software from social scientists, who use them to interrogate vast datasets on populations, historians and archaeologists are coming to use quantitative tools to model historic scenarios. The use of simulation modelling in history and archaeology is still regarded as something of a ‘dark art’ in many quarters. After all, the purpose of historical or archaeological enquiry is to reconstruct the past from evidence, be that evidence textual or material. Simulation, on the other hand, begins with evidential parameters and builds ‘could-be’ hypotheses, which are in some way credible, upon them. Such hypotheses cannot be tested or reproduced from other data, and as such — as methods of understanding the past — they are usually treated as a means to the end of interpretation, not an end in themselves.

One new project experimenting in this area is [ORBIS](#), designed by Walther Scheidel, a Roman historian, and developed by the technology team of Elijah Meeks, Karl Grossner, and Noemi Alvarez at Stanford

University. Put simply, the aim of ORBIS is to model the costs and times of travel between different points within the Roman Empire, over land or by sea or river. It does this by identifying factors such as wave height, terrain, mode of transport (e.g. camel train versus rapid military march) and mean climatic conditions. Using these, it computes how long it would take different categories of traveller using different means of transport, and at what cost, to traverse those distances by the optimal route. Results can be filtered by the fastest, cheapest, and shortest options. The site is free to use and, it should be noted, is a work in progress, with numerous upgrades, both technical and conceptual, promised for the future. Therefore, the observations that follow should be read not as criticisms, but as commentary on possible directions for its future development.

The site is structured with a simple and easy to navigate layer of content. Different tabs present an introduction and sections on ‘Understanding’, ‘Building’, ‘Using’, ‘Mapping’, and ‘Applying’ ORBIS, with various sub-tabs demonstrating particular aspects or works in progress on visualisation. The text and images of these are also available as a PDF download, which somehow gives credence to the authors’ claim that the site is not just a site, but also an online scholarly publication incorporating and building a software tool.

Usability aside, ORBIS is a new foray into rather perilous territory. The stated aim is ‘to understand the dynamics of the Roman imperial system as a whole.’ It explicitly eschews lower levels of detail and granularity, seeking rather to reconstruct, or rather model, systemic patterns of travel from pre-determined parameters. In its scope especially, ORBIS is distinct from many previous experiments in historical simulation, such as the [Road to Manzikert](#) project[1], or the [Battle of Trafalgar](#). [2] These projects (which are actually Agent-based Modelling exercises, and although ORBIS is not such a project it

nonetheless identifies ABM as a possible future direction of expansion) focus their simulations on distinct events, such as the progress of the Byzantine army to Constantinople in AD 1071, or the naval engagement of 1805; or on processes which are tightly constrained in time or space, with logical and easily-justified boundaries. ORBIS, with a set of environmental parameters across at least two climate zones, travel scenarios spanning three modern-day continents and some twenty modern-day national entities, and with few sharp lines around its edges in a qualitative historical sense, is therefore an undertaking of admirable, even daunting, ambition.

In any historical simulation project, the most important thing is to be clear about the limitations. In this, ORBIS cannot be faulted. The 'Building ORBIS' tab presents a long and detailed account of the principles guiding the selection of parameters for the modelling, and the evidence from which the quantitative aspects of the modelling are derived. This includes salient observations such as the fact that the average time computed for a ship's journey between any two ports will likely underestimate any actual time, since any ship is susceptible to a range of factors that might slow it down such as damage, piracy, or poor seamanship; whereas there are far fewer factors which could result in a sea journey being completed faster than would otherwise be anticipated by any predictable configuration (ancient or modern).

This set of quantitative predictions is constrained throughout by a focus on systemic features. The model includes only main arterial roads, rather than minor tracks, and excludes waterways that could have been navigated by small vessels or rafts. ORBIS keeps carefully to its turf, and avoids the temptation to be an oracle for any scholar wanting to construct any journey at any scale. It isn't, and it says it isn't. On the contrary, the scholarly referencing of ORBIS is impeccable, and the inclusion of well over 150 bibliographic references

further supports the authors' claim that this is a research tool and a scholarly publication combined.

The quantitative model that ORBIS relies on is explained in detail in the 'Building ORBIS' section, and its limitations carefully outlined. Secondary studies such as Pascal Arnaud's analysis of shipping are drawn on, as are pieces of primary evidence such as the tetrarchic price edict of AD 301, which is used to calculate the cost of maritime freight. These are used to justify the model quantitatively. The 751 locations are derived from the standard reference work for the ancient world's geography, the Barrington Atlas. A total of 84,631 kilometres of the Roman Imperial road network is mapped. The coastal sea-lanes mentioned in the historical sources are privileged in the ORBIS model. It is interesting that in a previous iteration of the model, the project considered using a minimum range approach, whereby connections between coastal sites within a minimum radius (of 500 miles) would have been privileged — the dismissal of a purely abstract quantitative approach in favour of one derived from sources is, I think, an interesting reflection of how the humanities' needs for quantitative methods differs from those of the social sciences.

I look forward to the promised future upgrade where the user will be able to view information on seasonal hazards into the outcomes of the calculations they have submitted. River transportation is, in general, far more difficult to reconstruct, due to the shifting character of the landscape, modern human intervention in Europe's waterways, and paucity of documentation. Accordingly, ORBIS's model relies on comparanda from the medieval period. This of course introduces additional levels of complexity and uncertainty, and one area that must be developed further is the question of how this greater cumulative fuzziness might affect individual applications of the model.

So: the constraints of the model are described in detail, and ORBIS is scrupulously careful not to make unfounded claims. However, the constraints themselves occupy a place in the debate around quantitative methods in the reconstruction of the past, and future iterations of the project must — in my own view — continue to exercise caution against the temptation to lapse into environmental determinism. Future iterations must also continue to make clear that ORBIS does not try to simulate agency, and that this requires the user to use it only in the context of their own analysis or interpretation, and not as an independent generator of research conclusions. For example, ORBIS states that ‘while much of the sea is in theory navigable without major restrictions, sailors would often follow established routes, and certain roads and rivers were more heavily used than others and hence more vital to the functioning of the system.’<sup>[3]</sup> This seems to assume that the only constraints followed by sailors navigating water was wind and currents, and ‘black boxes’ the reasons why certain routes became ‘established’ and, crucially, how these factors might change over time. One can — as I do — fully accept ORBIS’s premise that ‘No one traveler would encounter such [simulated] outcomes except by chance,’ and appreciate that the processes being simulated are pan-Imperial and systemic.<sup>[4]</sup> However, when one considers the great events that shaped the Roman world over time — wars, assassinations, expansions, incursions — the limitations of adopting a purely abstract approach based on systemic parameters, become even more apparent. This tantalizes the user with the possibilities that, one hopes, future iterations of the tool will explore. For example, it is difficult to see how the processes we can discern from non-simulative historical sources, such as the consolidation of the Roman Empire’s frontiers between AD 140 and AD 160 could be explored using a tool whose only constraints are environmental. One could speculate that this is merely one factor of a similarly high/systemic level, specific to the Roman period, which

the user could include. As the project develops, the ‘Applying ORBIS’ tab, which presents instances of the model’s use in academic work, must expand and document the tool’s use in these areas.

My own experiment with the ‘Mapping ORBIS’ section of the site, where one can try the tool, concerned a relatively short-haul route in northern Britannia between Segedunum (Wallsend, Newcastle) and Luguvalium (Carlisle) — about five sixths of the 74 mile expanse of Hadrian’s Wall — mainly because I have walked this route myself (or at least the sections of it followed by the modern-day National Trail). ORBIS computes the ‘cheapest’ option as an entirely plausible 4.7 days. However, this includes river transportation as far as Corstopitum (Corbridge), and whether the Tyne was navigable that far upriver in this period is a far from straightforward question, underlining the uncertainties raised about ancient river conditions. When one seeks to exclude river transportation from the calculation, an error message is returned. It seems not to be possible to specify travel overland from Segedunum to, for example, Cilurnum on the Wall itself, and then link up with the overland route from thence to Luguvalium. While one takes the point that a section of river travel is needed for the journey to be literally fastest, cheapest, or shortest, it is surely important for the user to be able to include only modes of transport that they know they are interested in. Otherwise it is the model, not they, who is in the driver's seat. The experience itself however is user-friendly and straightforward, and one can easily see how it could be usefully combined with the ‘Network Visualisation’ tool, which allows you to see and navigate the connections between the different points, under the ‘Understanding ORBIS’ tab.

It is a personal belief of mine that the concepts of ‘quantitative’ and ‘qualitative’ data are direct imports from the social sciences, and we in the humanities do not always think in enough detail about what this

distinction means for humanities material. ORBIS is an exciting and innovative experiment in simulation modelling in history and archaeology, which starts to critique this distinction; even if, by its own clear admission, it does not yet have the answers. This project is only just starting out, and it has made a convincing pitch for its approach to what we might think of as a qualified quantitative network and spatial visualisation of a generalised view of travel in the Roman world. The possibilities for the future are immense, and one can only begin to imagine the strides that could be made were it to follow up on the aim stated in the documentation to build links with the [Pleiades](#) and [Pelagios](#) projects. I, and I am sure many others, will be watching ORBIS's progress with great interest.

#### Notes:

[1] Vince Gaffney, Georgios Theodoropoulos, Rob Minson, and Phil Murgatroyd, "The Road to Manzikert," *Digital Classicist*, 2009. <http://www.digitalclassicist.org/wip/wip2009-02pm.pdf>.

[2] Giuseppe Trautteur and Raniero Virgilio, "An agent-based computational model for the Battle of Trafalgar: a comparison between analytical and simulative methods of research," *Proceedings of the Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2003. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1231440>.

[3] Understanding ORBIS tab.

[4] Ibid.

# The French Book Trade in Enlightenment Europe

## PROJECT OVERVIEW

### **The French Book Trade in Enlightenment Europe Project and the STN Database**

Simon Burrows & Mark Curran . . . . . 53

## CASE STUDY

### **How Swiss was the Société Typographique de Neuchâtel? A Digital Case Study of French Book Trade Networks**

Simon Burrows & Mark Curran . . . . . 56

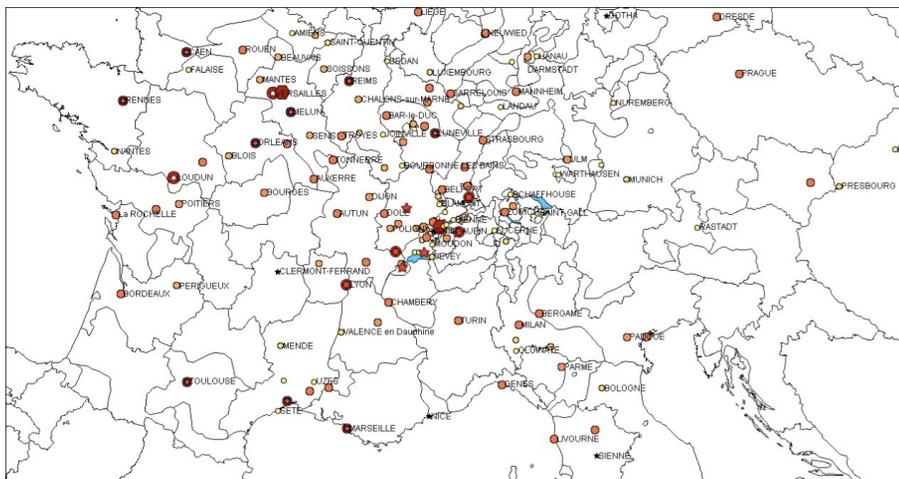
## REVIEW

### **Review of The French Book Trade in Enlightenment Europe, 1769-1794: Mapping the Trade of the Société Typographique de Neuchâtel**

Sean Takats . . . . . 66

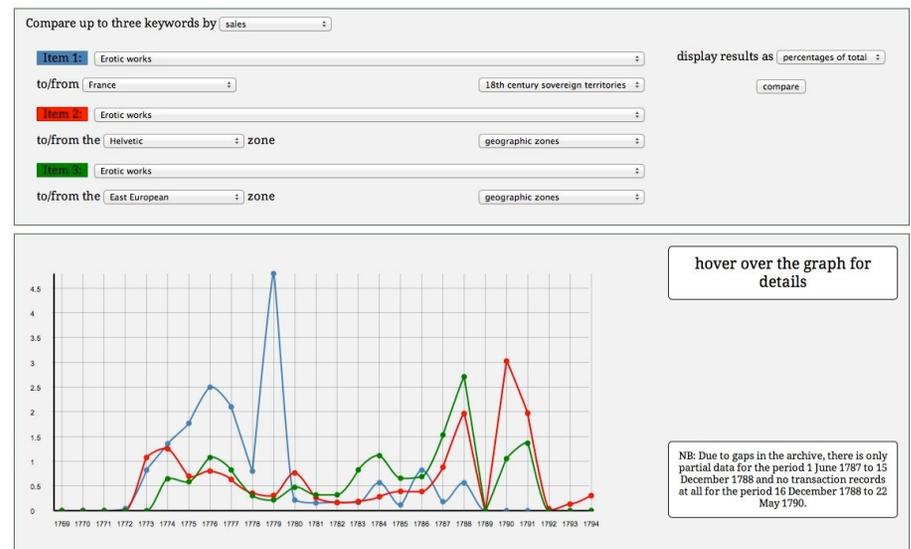
# The French Book Trade in Enlightenment Europe Project and the STN Database

The French Book Trade in Enlightenment Europe (FBTEE) database maps the trade of the Société Typographique de Neuchâtel (STN), a celebrated Swiss publishing house that operated between 1769 and 1794. The database was published online on June 25, 2012 and is available [here](#). It is intended to be a major new tool for literary and historical research in Enlightenment studies, book and business history, and bibliography.[1]



STN Sales to France and Central Europe

As the STN sold the works of other publishers alongside its own editions, [the STN archives](#) have long been considered a representative, as well as uniquely rich, source for studying the book trade and dissemination of ideas in the late Enlightenment. Moreover, because the STN traded from the Prussian ruled Swiss-principality of Neuchâtel, it could deal in all genres of book, supplying clandestine copies of illegal and pirate editions banned in France and many other states. Strategically based in the middle of Europe and shielded by the Prussian monarchy, the Neuchâtelois booksellers developed client networks spanning the continent from Dublin to Naples, Lisbon to St Petersburg. Thus the STN data can offer internationally comparative insights into reading tastes during the late Enlightenment.



Comparative STN 'Sales' Trends for Erotic Works in France (Blue), Switzerland (Red), and Eastern Europe (Green)

The database was constructed using the STN's double-entry account books. They make it possible to reconstruct almost the whole of the STN's trade – recording both from whence every book came and to

whom it was sold, as well as when each transaction occurred. Alongside this information, the database records the genre, function, content, and legal status of each work sold.



Arts & Humanities  
Research Council



UNIVERSITY OF LEEDS

## The STN Interface

There are many ways into the STN database. Its web-based user-interface allows users to interrogate the database by a considerable range of fields, including author, title, publisher, subject matter, genre, and time-period in addition to the names, professions, and places of residence of individuals who supplied or ordered STN books. Specialised queries allow users to go further, for example identifying [bestselling texts](#) or [authors](#) for any given place and time period. Using these tools, it is also possible to chart reading tastes

across Europe; changing patterns of demand over time; and networks of exchange in the print-trade. For more sophisticated queries, there are a series of [option menus](#) or filters which allow users to interrogate subsets of the data (e.g. sales of illegal books; the trade of female book dealers; or the dissemination of English works in French translation).

	total number	as % for that period
1. Dissertation sur l'établissement de l'Abbaye de S. Claude	1,562	11.09
2. Ecole (!) du bonheur	1,090	7.74
3. Compte-rendu au Roi	764	5.43
4. Essai sur le préjugé subsistant contre les familles des condamnés	703	4.99
5. Dictionnaire portatif de la campagne	700	4.97
6. Mémoire donné au roi par M. Necker en 1778	430	3.05
7. Lettre du marquis de Caraccioli à M. d'Alembert	400	2.84
8. Requête au roi sur la retraite de Mr. Necker	400	2.84
9. Réponse du sieur Bourbonlon... au Compte rendu au roi par M. Necker	400	2.84
10. Commentaires (les)	400	2.84

## Top 10 Works Going to Besançon

The authors of the database believe that their meticulously designed data structures, tools, and interface are ideally constituted for other studies of the production, distribution and dissemination, and reception of printed works and other cultural artefacts. They hope to see these features widely adopted, in order to facilitate other studies and the emergence of a network of interoperable projects.

5th June 1782	In: 25 full copies of Jonathan Wild le Grand from Nouffer/Nouffer & Bassompierre/Nouffer de Rodon (GENÈVE)	<a href="#">show/hide details</a>
11th July 1782	Out: 6 full copies of Jonathan Wild le Grand to Poincot (VERSAILLES)	<a href="#">show/hide details</a>
7th August 1782	Out: 6 full copies of Jonathan Wild le Grand to Rüdiger, Christian (MOSCOU)	<a href="#">show/hide details</a>
7th October 1782	Out: 2 full copies of Jonathan Wild le Grand to Girardet, Samuel (LOCLE, Le)	<a href="#">show/hide details</a>
21st October 1782	Out: 2 full copies of Jonathan Wild le Grand to Comte (BELFORT)	<a href="#">show/hide details</a>
30th October 1782	Out: 1 full copy of Jonathan Wild le Grand to Roeder, Jean-Charles (YVERDON)	<a href="#">show/hide details</a>
31st December 1782	Bilan: 9 full copies of Jonathan Wild le Grand	<a href="#">show/hide details</a>
6th February 1783	Out: 1 full copy of Jonathan Wild le Grand to Bouvet (TROYES)	<a href="#">show/hide details</a>
7th June 1783	In: 27 full copies of Jonathan Wild le Grand from Nouffer/Nouffer & Bassompierre/Nouffer de Rodon (GENÈVE)	<a href="#">show/hide details</a>
14th June 1783	Out: 4 full copies of Jonathan Wild le Grand to Chirol, Barthelemy (GENÈVE)	<a href="#">show/hide details</a>
12th July 1783	Out: 5 full copies of Jonathan Wild le Grand to Rüdiger, Christian (MOSCOU)	<a href="#">show/hide details</a>
12th July 1783	Out: 3 full copies of Jonathan Wild le Grand to Muller, Charles Guillaume (SAINT-PETERSBOURG)	<a href="#">show/hide details</a>

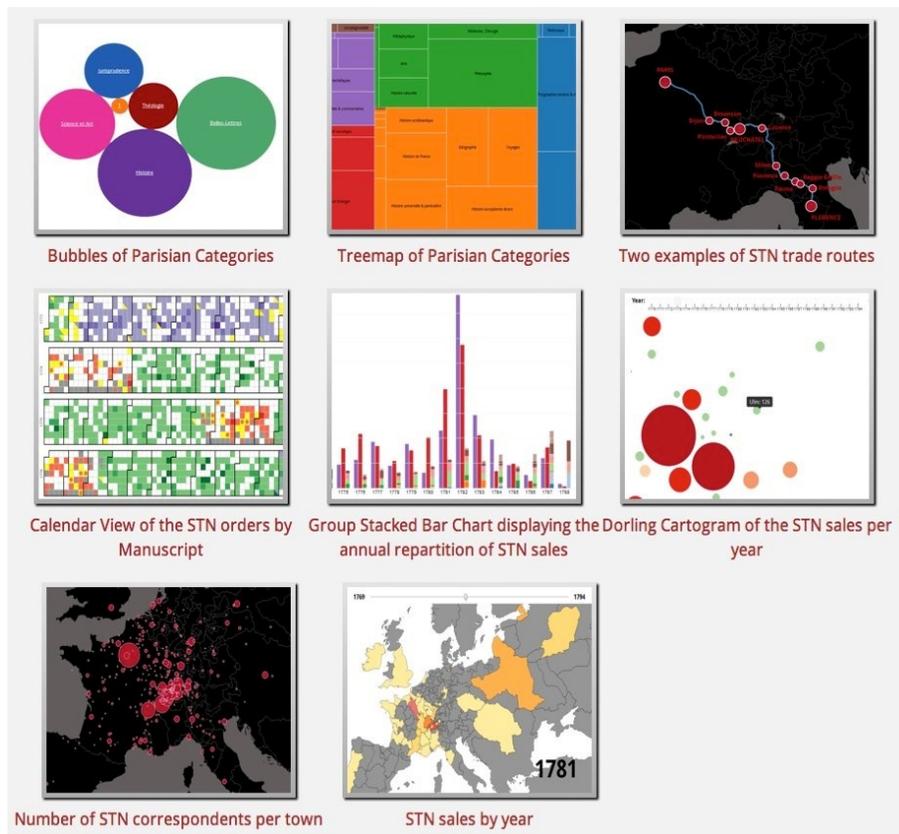
## Building Blocks of the Database: Event Data for Fielding's Jonathan Wild Le Grand

The database resource offers a range of supporting materials to help users explore, understand, and mine the data. A drop-down 'Help and Resources' menu inside the data interface contains a [User Guide](#) and various other tools to help users orientate themselves ([Designer notes](#), lists of [Abbreviations](#) and [Manuscript Sources](#), [Database totals](#), [Visualisations](#) and [Videos](#), [Designer Credits](#) etc.). The [video instructional and walk through materials](#) are also available through the project website. The website also gives an [overview of the project](#) more generally and some of the [uses of the database](#). Latest news, updates, and links to resources are available at the official FBTEE [project blog](#).

Interpretative insights into the project's findings are available in a [series of podcasts](#) and [publications](#). In the case study that follows, the database's authors showcase some of its main uses. In the process they explore an overarching interpretative question that both shapes and bedevils the interpretative volumes they are currently writing on the structure of the European book trade (vol. I) and the dissemination of ideas and discourses in the late Enlightenment (vol. II).

**Notes:**

[1] The French Book Trade in Enlightenment Europe project was funded by a £355,485 grant awarded to Simon Burrows by the British Arts and Humanities Research Council (AHRC). The authors are grateful to the AHRC for making their work possible, and to the University of Leeds for supplementary support and for hosting the project.



The FBTEE Project Visualisation Gallery

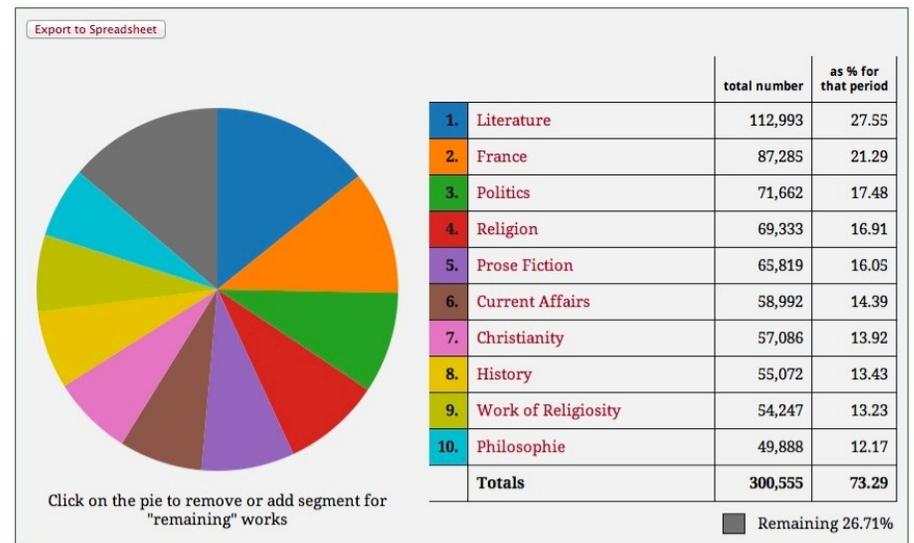
SIMON BURROWS & MARK CURRAN

# How Swiss was the Société Typographique de Neuchâtel? A Digital Case Study of French Book Trade Networks

Should the famous Société typographique de Neuchâtel (1769-1794) (STN) be viewed as a typical and representative ‘European’ publisher-bookseller, a ‘print shop across the border’ that offered the latest product of the mud-raking underground to the French market, or a provincial, peripheral, and Swiss-focused distraction? The question needs revisiting before we both trumpet the value of our recently published French Book Trade in Enlightenment Europe (FBTEE) database — which recreates the society’s entire knowable bookselling business — and assess its potential for substantially revising scholarly understanding of the eighteenth-century French book trade and, dare we dream, the enlightenment.[1]

Fortunately the database itself is ideally structured for providing powerful new geographically-based insights, since its fundamental purpose is to pinpoint and map transfers of books between the STN and its clients in time and space. In addition to offering carefully structured taxonomic and bibliographic data on books — capable of dealing with issues such as false imprints, original language of publication, and the legal status of editions — the database offers multiple forms of spatially related data. It records where clients were

based, where books were published and from whence they were despatched to the STN, as well as the languages in which works were originally published. For purposes of analysis, it also groups geographic locations into provinces, countries, and wider regions, as well as into a variety of transnational entities (e.g. ecclesiastical lands; the Holy Roman Empire; university towns). Likewise it groups clients by the type of trade they conducted with the STN and their stated professions, as well as into wider economically and professionally-defined groupings.



The Enlightenment Captured? Top Ten Subject Keywords By Sales In The STN Database.

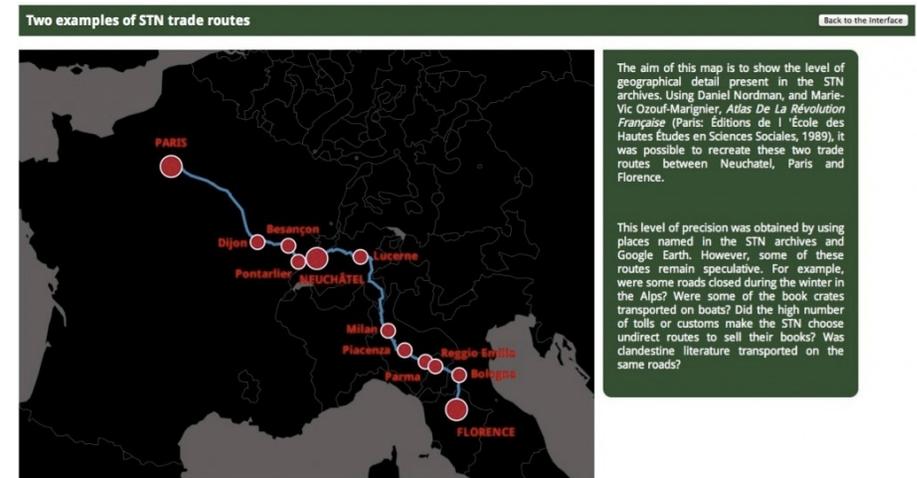
By using the project’s on-line interface, it is thus possible to rank, compare, or map clients, books, authors, or places by volume of trade over time, using increments as small as a single day. This makes it possible to reveal instantly, in ways not feasible in pre-digital studies, both macro-sales trends or bestseller listings and more finely calibrated micro-histories of particular books, places, or clients. In

order to understand how far the STN should be considered a distinctively Swiss venture, we will need to look at all three, using both macro- and micro-lenses.

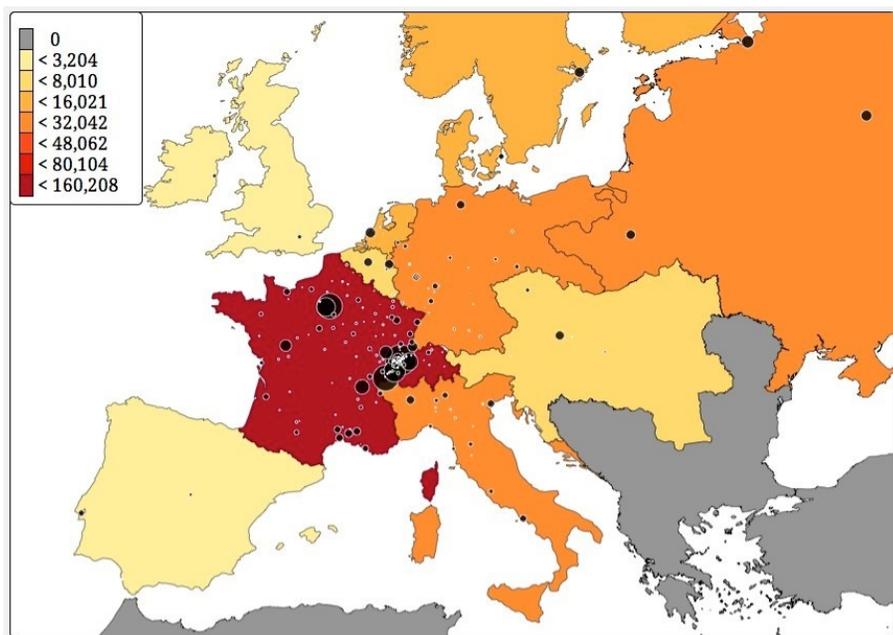
Robert Darnton, of course, has already opined on the very same question in his seminal *The Forbidden Bestsellers of Pre-Revolutionary France*.<sup>[2]</sup> Employing an elegant argument that centred upon the idea that European printers located from Amsterdam to Geneva swapped substantial volumes of new works amongst themselves as soon as they were printed, Darnton constructed the STN as a representative business capable of supplying more-or-less anything.<sup>[3]</sup> The society's sales patterns, it was suggested, differed little from those of other extra-territorial publishers selling clandestine works into France from warehouses in Switzerland, Holland, Belgium, Bouillon, or the German Rhineland. This thesis proved so persuasive that it has become the main prism through which the STN, and indeed the entire late *ancien régime* publishing trade, is generally understood.<sup>[4]</sup> Indeed, its stylish simplicity helped us to justify our own project to our Arts and Humanities Research Council (AHRC) funders: a database of a single publisher's trade with wider representative value would allow scholars to expansively compare European reading tastes on the eve of the French Revolution. Who could resist green-lighting such a tantalising proposal?

Better understanding the more complex realities of the STN's market orientation through the FBTEE database must begin with a widescreen view of the 400,000-plus books that the Neuchâtelois dispatched to clients across Europe. Ranking the broadest-available geographic entities in the interface (as on the map below or – for the interactive version – [here](#)) reveals that Helvetic traders – that is to say those based in the geographic region of Switzerland, including states that were outside the eighteenth century Swiss Confederation such as

Geneva and Neuchâtel – were the biggest recipients of STN-traded books. They took 39% of the society's unit stocks. French traders received a little under 37%. Wider European clients combined accounted for just over 24%. Caveats, as always, are necessary before making too much of these results. Many books were exchanged locally rather than 'sold'; entire editions were regularly commissioned; newly printed works were dispatched to established clients speculatively; and a few copies went to censors, the authorities, or political allies. Moreover, countless works shipped to Geneva or Lausanne clearly went further afield. None the less, the sheer volume of data at least compellingly shows that the STN was much more than a simple 'print shop' across the French border.



Two Examples Of STN Trade Routes

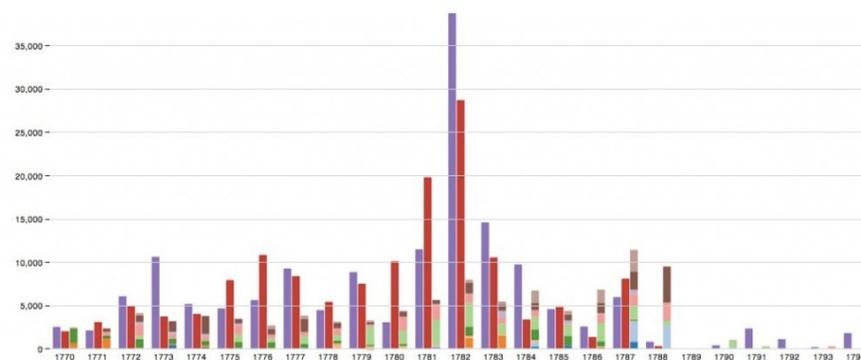


STN Unit Sales By Geographic Region, 1769-1794

Pinning down what sort of operation they *were*, however, is complicated by significant temporal fluctuations in the data. At the core of the FBTEE database sit over 70,000 ‘transaction’ or ‘event’ entries, that record day-and-date when the STN’s orders were sent and received (or, more accurately, when these events were accounted for). A calendar of the society’s entire trade can thus be recreated and is visualised [here](#). Equally, we might try to examine the STN’s shipping records on a daily, weekly, monthly, or annual basis, for individual authors, places or, as in the example given [here](#), particular titles. Our colleague Vincent Hiribarren’s interactive visualisation of trade on an annualised basis, found in our [‘visualisation gallery’](#) and below, most clearly shows how the society’s market orientation changed over time.<sup>[5]</sup>

Between 1770-1772, as the Neuchâtelois built up their business, their client base largely remained local or based along the most accessible communications routes along the Rhine and Rhone corridors. By the mid-1770s they had established extensive networks across Europe, but particularly in France, which remained their largest market until the early 1780s. This was also the period during which the STN’s trade was at its height (see graph below or its interactive version [here](#)). However, as Leeds PhD student Louise Seaward is currently establishing, from 1783 French measures aimed at stopping the illegal book trade added to the costs and risks of doing business in France.<sup>[6]</sup> Consequently, the French trade declined and the society began reorientating their trade towards Italy and Northern and Eastern Europe.

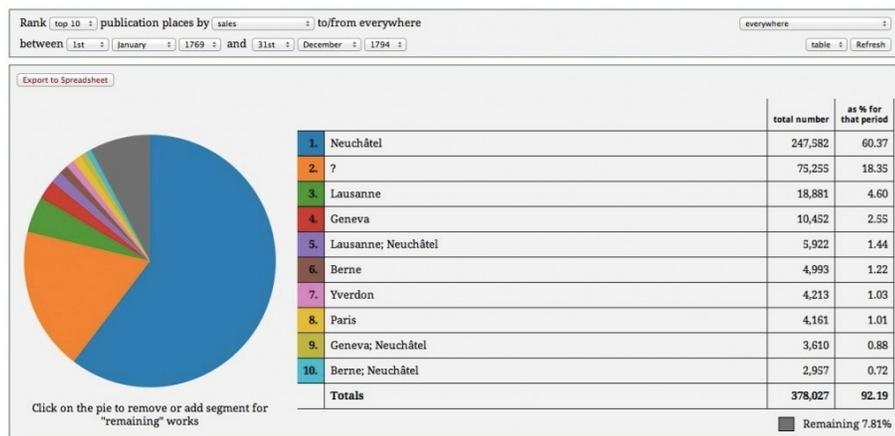
This Group Stacked Bar Chart displays the annual repartition of STN sales in three different geographic zones (Swiss, French and others). The aim of this chart is to compare the importance of the Swiss and French markets from 1769 to 1794. Noticeably, other markets became more prominent when sales to France dropped in the 1780s.



Annualised STN Sales By Geographic Zone. (NB. Red = France; Mauve = Switzerland)

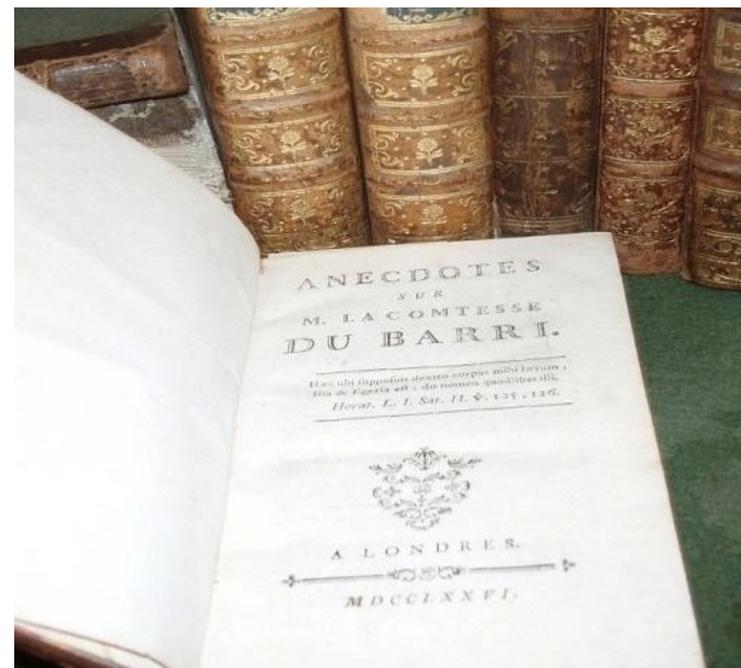
As their market orientation shifted in accordance with changing economic and regulatory sands, however, one important thing remained constant: the society predominantly sold Swiss books. 60% of unit shipments were of Neuchâtel editions, largely those from the

society's own presses. The majority of the remaining top twenty places from which the society sourced works were also Swiss. In total, over 75% of STN 'sales' (313,089 units) were for editions positively identified as having been published in Switzerland. By contrast, only one percent of unit sales appear to have involved Parisian editions; while just over half of one percent were from London (see the table and graph below or its interactive version [here](#)). Of course, some of the recorded imprints may have been false, although where possible this has been taken into account — this search runs on the imprint stated on the title page unless corrected by bibliographic information from library catalogues. But a bigger uncertainty comes from the fact that almost one-fifth of STN *envois* (76,544 units) relate to bibliographically unresolved works or those that carried no imprint *at all*. Many of these works were clandestine or illegal in some way; their publishers wishing to conceal or falsify their place of origin for marketing purposes or to protect their backs. How can we know where these came from?



Most Common Places Of Publication By STN Unit Sales

A case-by-case analysis can be very revealing. It can, for example, help to expose the likely publishers of specific clandestine editions. Mathieu-François Pidansat de Mairobert's underground scandalous biography of Louis XV's final mistress, the *Anecdotes sur Madame la Comtesse du Barry*, was first published in 1775. The extraordinarily high numbers of this work that the STN received from François Grasset of Lausanne in 1776 and early 1777 suggest that he was the publisher of the 1776 Swiss edition. Further, the fact that the STN took a remarkable 825 copies from Grasset during this period suggests that, in an era when a normal print run was 500-1,000 copies, Grasset's print run was perhaps a phenomenal 5,000 or more (for further details [see here](#)). Consumers could not resist political and sexual scandal.

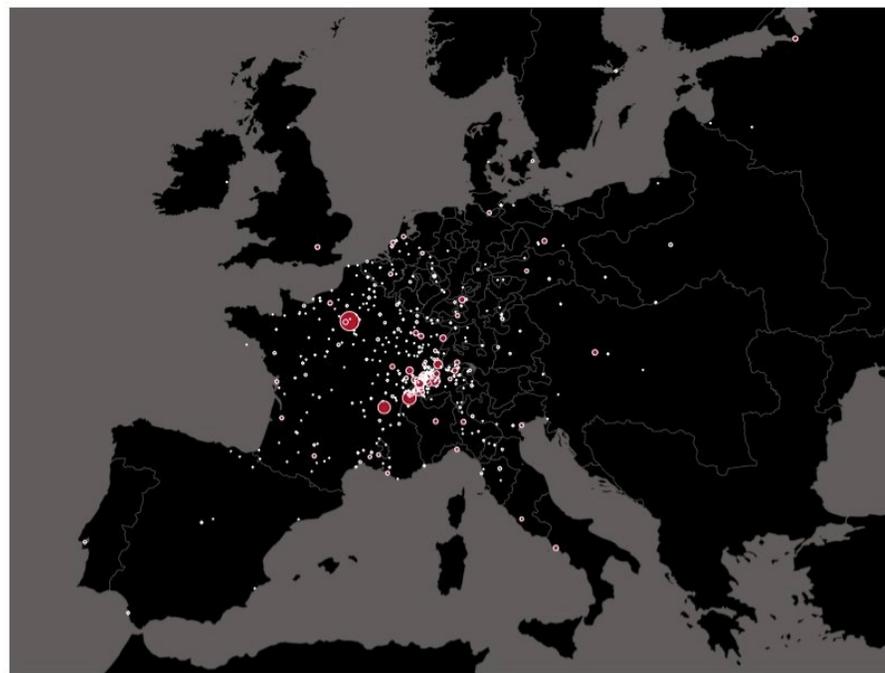


Mairobert's Best-Selling *Anecdotes Sur Madame La Comtesse Du Barri*

But ploughing through nearly 4,000 distinct bibliographical entities whilst trying to keep the big picture in mind is an exhausting and impractical undertaking. As a consequence, the FBTEE database interface has been designed to make it easy to seek the origins of such works on a more general level. A number of filters have been incorporated that allow users to interrogate pre-determined subsets of the data. In all there are eight options menus (with summaries [available here](#)), that can be used to delicately but powerfully nuance global dataset queries. Searches can be restricted to works which were illegal or in translation; to female clients; to wholesale book dealers; or to data drawn from various types of sources. Most importantly for current purposes, we can limit queries to certain types of edition, as defined by publishing origins. If we select works whose place of publication is unknown, then rank the supply origins of such works by geographic zone [as has been done here](#), we learn that at least 56% of such works were sent to the society from the Helvetic zone. From what we already know about the STN's supply patterns, we can therefore confidently suggest that somewhere between 90% and 95% of the books that they traded were Swiss editions.

At this point, all hopes of seeing the STN as a representative European publisher-bookseller depend upon a simple question: to what extent did Swiss books mirror those printed elsewhere? The reconstruction of the STN's intellectual milieu through an examination of their correspondence networks allows for some initial insight here. A cartographic visualisation of the locations of the STN's correspondents (see below and [here](#)) shows, as we might expect, that clients were concentrated in Switzerland, Paris, frontier regions of France, and along the Rhine, Rhone, and other prominent river corridors in France, Germany, and Italy. However, the STN's general correspondents were comprised of everyone from booksellers and publishers to Versailles

politicians such as the comte de Vergennes, the French foreign minister who kept a watchful eye on the extra-territorial book trade. They include the chandlers, printfounders, inkmakers, and paper merchants who supplied printshop essentials, and business agents, postmasters, and wagoners who facilitated the movement of books around Europe.<sup>[7]</sup>

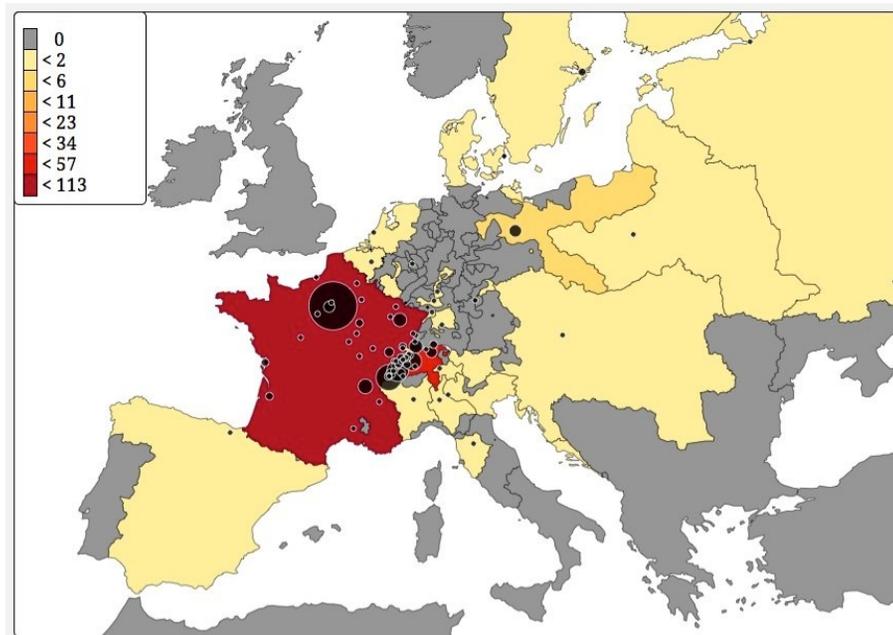


Map Of STN Correspondents By Place Of Residence

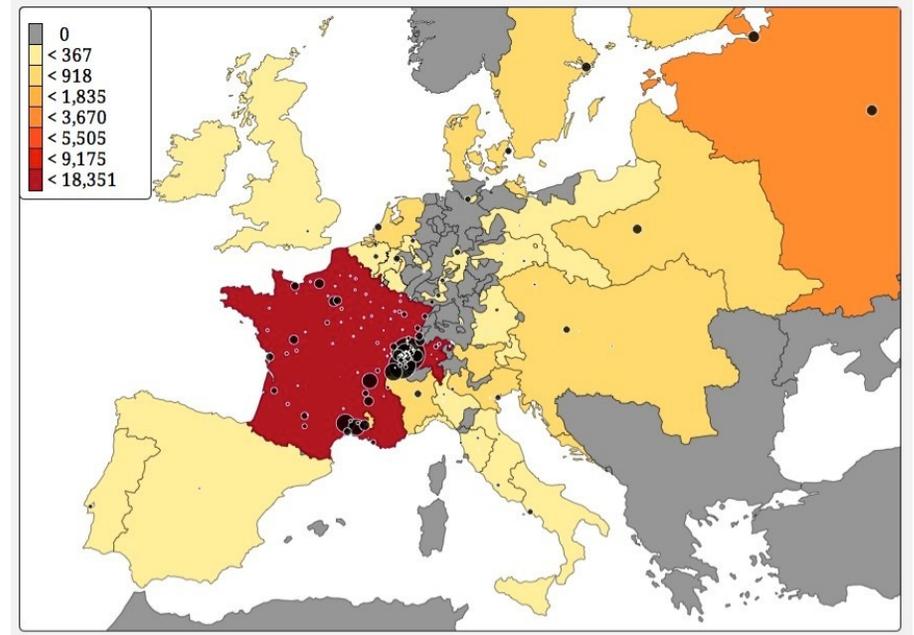
Once (as below and [here](#)) we restrict the focus of enquiry to 'Men of Letters' — the intellectuals and academics who supplied the society with fresh manuscripts to publish, or who wrote to Neuchâtel with the latest literary intelligence — a much more clustered grouping emerges. The vast majority of such individuals were based either in France — over two-thirds of them lived in Paris — or in geographic Switzerland.

Again, some caution is necessary: the database counts the number of times that a professional designation is used, and since some individuals have more than one professional designation, a little double-counting is unavoidable. In the case of this search, thankfully, there is an overlap in only 5% of cases (for full data see [here](#)). Once again, the results are sufficiently overwhelming to engender confidence. With 112 and 95 correspondent designations respectively, France and Switzerland were home to 88% of the 'Men of Letters' who corresponded with the STN. Most of the 12% based elsewhere wrote from courtly capitals in Germany, Italy, and Northern Europe, from Stockholm, Warsaw, Vienna, St. Petersburg, and above all, with five designations (but only four correspondents), Berlin, home to the Prussian monarchs, who were then sovereign princes of Neuchâtel.

Thus, a preliminary sketch suggests that the STN's intellectual world was more Franco-Swiss than the raw bibliometrics reveal. This impression is strengthened when we [browse outlying centres such as Berlin](#) to see who the 'Men of Letters' in residence there actually were. This query reveals them to have been the celebrated writer and bookseller Friedrich Nicolai; the bestselling *philosophe* and slave trade abolitionist, the *abbé* Guillaume-François-Thomas Raynal; the well-known Italian mathematician and astronomer, Giovanni Francesco Mauro Melchiorre Salvemini di Castiglione (who was double-counted); and the physician Nicolas de Béguelin, who was a tutor to Frederick the Great and became Director of the Berlin Academy. Of these, Béguelin was Swiss, Raynal French, and Castiglione a long-term resident of Lausanne, where he embraced Calvinism and taught for 15 years.

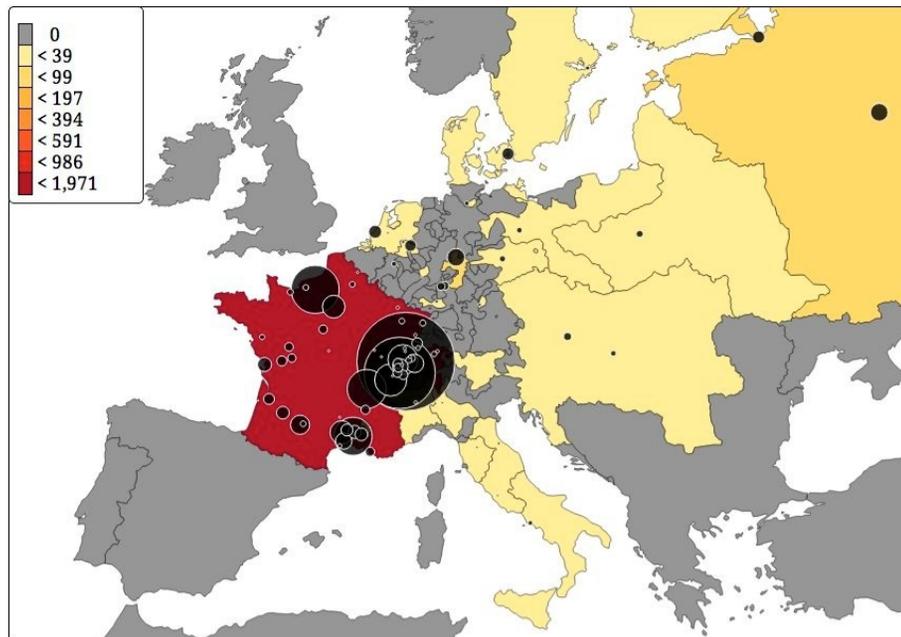


Places Of Residence Of 'Men Of Letters' Who Corresponded With The STN



STN Sales Of 'Works Of Religiosity'

This analysis can be pushed further by exploiting the FBTEE database’s keyword system for describing the content, genre, and ideological tendencies of books. Certainly there are areas where the Swiss Protestant ethos comes across strongly. Unsurprisingly, this is particularly true of the religious books the STN traded: the vast majority were either Protestant or generically Christian. When, as above, we [map ‘Works of Religiosity’](#) (our term for works written with a clear Christian purpose), the Protestant arc is clearly visible, running from French-speaking Switzerland down to the Southern French strongholds such as Nîmes, then up through Montauban and Bordeaux to La Rochelle. The trend is clearer still if we map the dissemination of Bibles by the STN.

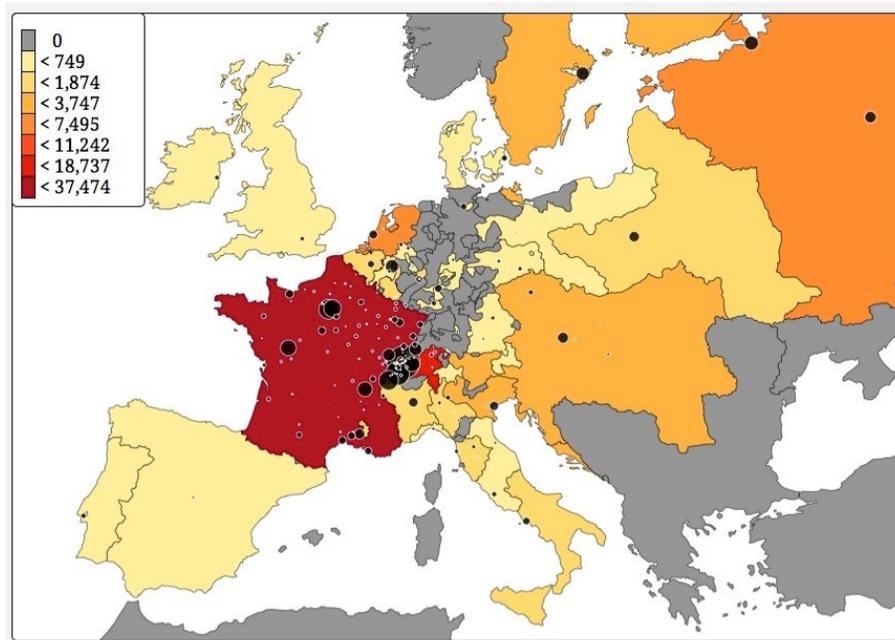


Dissemination Of Bibles By The STN, 1769-1794

The religious works sold by the STN were clearly catering to its Protestant co-religionists. This view is also reinforced if we use the database to compare sales of books to the [Protestant](#) and [Catholic](#) clergy. The Protestants focused on religious texts for themselves and their flocks; the Catholics — who to be fair included many worldwide literary *abbés* — certainly did not. Perhaps more surprisingly, [the scientific works traded by the STN](#) appear to have been distinctively Swiss. Leaving aside textbooks, commissioned works, reference books, and works touching upon science incidentally, it is Swiss luminaries such as Charles Bonnet, Albrecht Haller, and Samuel Tissot who dominate their scientific bestseller list.

Yet, at this stage at least, it would be inappropriate to take the argument too far. The Swiss booksellers were significant cultural intermediaries, facilitators of significant cross-cultural transfer.<sup>[8]</sup> They published their own translated editions of foreign works, especially English and German literary works, and sold them across francophone Europe (see map below and [here](#)). In all, the STN sold 300 works translated from English or German, which accounted for almost 10% of total shipments (40,050 out of 411,000). Equally, the society sold the literature of travel and exploration in large quantities (23,975 units), in the process familiarising European readers with other European and world cultures. We might also note [here](#) the large numbers of works addressing historical themes, including textbooks, biographies, and enormous multi-volume national histories. Thus, while in some domains the stock of the STN seems to have been particularly Swiss in subject matter, in others it was extremely cosmopolitan. Moreover, as we first outlined in a lecture at the Library of Congress (available as a podcast [here](#)), the STN seems to have sourced books on particular themes and in specific genres in broadly — and often remarkably — similar proportions from both within and without Switzerland. This

suggests that Swiss preoccupations were very similar to those of other (francophone) Europeans. On the level of generic subject matter, if not specific orientation, the STN's stock appears to have a wider representative value.



### Dissemination Of English And German Works By The STN

So to the question 'how Swiss was the STN?' this brief case study offers a rather complex answer. Certainly, any future treatment cannot ignore the fact that the STN's stock was Swiss, and its market orientation was less towards France than has often been thought. Ideologically, religiously, politically, and, perhaps more surprisingly, scientifically it had strong Swiss biases, too. But at the same time, in many domains, the STN shared in and promoted the intellectual cosmopolitanism of the enlightenment and its common concerns. From a distinctly Swiss base, it was outward facing and firmly embedded in the literary culture

of its day. The task of eliciting and explaining what this means for our understanding of the book trade in particular and the enlightenment in general must, however, be left to another day.[9]

Meanwhile, the further development of the project points us in multiple directions. The simplest means to expand our work would be to create databases of other comparable archives — e.g. the archives of the Luchtman publishing company in Leiden or the registers of the Veuve Desaint in Paris, both of which are contemporaneous with the STN. This might give further insights into the distinctively Swiss qualities of the STN. More valuable to the wider scholarly community, however, are plans to adapt the data structures and data entry tools to facilitate projects drawing on other bibliometric sources to map and measure the dissemination and reception of texts. Libraries and archives are teeming with such sources, which include publishers records; catalogues of public and private libraries and book dealers; confiscation and publishing permission registers; contemporary book reviews; book advertisements; comments in private correspondence; indices of banned books, etc.

The FBTEE database already contains data drawn from several such sources, but to better integrate them, it needs to structure all the 'events' they record in a similar way to our book trade transactions. This would, in theory, make it possible to map, list, or interrogate all types of event at once, and to generate a unitary and rich statistical reception history of books, in much the same way as Amazon.com presents sales statistics and reviews on contemporary merchandise. It also offers the possibility of gathering more qualitative data, not dissimilar perhaps to that in the various national [Reading Experience Databases](#).

Finally, our tools and approaches can be applied to projects focused on other times and places, although they are probably best suited to societies where access to print and print technologies were limited enough to make such research manageable. The cut off point in Western Europe or North America is thus the second printing revolution of the early to mid-nineteenth century; in colonial Australasia it is probably somewhat later; and in parts of Africa more recent still. The ultimate prize, then, is a unitary but flexible digital system for researching, accessing, and calibrating the dissemination and influence of ideas, knowledge, and culture across the entire print era. The FBTEE database is a small but foundational step in that direction. In this sense, at the very least, our single Swiss publisher's archive has proved to have a more global representative value.

#### Notes:

[1] This article and the database on which it draws are outcomes of the Arts and Humanities Research Council-funded project on the French Book Trade in Enlightenment Europe, headed by Simon Burrows at the University of Leeds. The database is co-authored by Simon Burrows and Mark Curran, who wish to thank the AHRC, the University of Leeds and project team members Sarah Kattau, Amyas (Henry) Merivale, and Vincent Hiribarren for their support throughout the project.

[2] Robert Darnton, *The Forbidden Best-Sellers of Pre-Revolutionary France* (London and New York: Norton, 1996). For his statistical appendices, which inspired the STN database, see Robert Darnton, *The Corpus of Clandestine Literature in France, 1769-1789* (London and New York: Norton, 1995).

[3] Darnton's argument is deconstructed in Mark Curran, "Beyond the Forbidden Best-Sellers of Pre-Revolutionary France," *Historical Journal* (forthcoming). A more extensive examination and re-evaluation of the book trade will appear in his monograph from this project (see note 9 below). For an overview of the STN's illegal trade across Europe in the light of the database see Simon Burrows, "French Banned Books in International Perspective, 1770-1789" in *Experiencing the French Revolution*, ed. David Andress, forthcoming SVEC (Oxford: Voltaire Foundation, forthcoming).

[4] Besides the works already noted, the major published studies of STN and its trade are *L'Edition neuchâteloise au siècle des Lumières: la Société typographique de Neuchâtel (1769-1789)*, ed. Michel Schlup (Neuchâtel: B.P.U.N., 2002); *Le Rayonnement d'une maison d'édition dans l'Europe des Lumières: la Société typographique de Neuchâtel 1769-1789*, ed. Robert Darnton and Michel Schlup (Neuchâtel: B.P.U.N.; Hauterive: Editions Gilles Stinger, 2005); Jeffrey Freedman, *Books Without Borders in Enlightenment Europe: French Cosmopolitanism and German Literary Markets* (Philadelphia, PA: University of Pennsylvania Press, 2012).

[5] It is appropriate here to acknowledge Vincent Hiribarren's sterling work at developing maps and visualisations for the project, initially on a shoestring budget. All the visualisations from the project used in this paper are the result of his work. For more of Dr. Hiribarren's work see <http://www.vincenthiribarren.com/>.

[6] Louise Seaward is currently in the final year of her doctoral project on French policing of the extra-territorial book trade, and has worked closely with the project team.

[7] A handlist of most of these people is available on-line [here](#). It lists correspondents and their professions under a single place of residence / correspondence. The list has been used for previous explorations of the STN's trade networks by Frédéric Inderwildi and Thierry Rigogne. See Frédéric Inderwildi, "Géographie des correspondants de libraires dans la deuxième moitié du 18e siècle: La Société typographique de Neuchâtel, Cramer et Gosse à Genève," *Dix-huitième siècle* 40 (2008): 503-22 ; Thierry Rigogne, "Librairie et réseaux commerciaux du livre en France à la fin de l'Ancien Régime" in *Le Rayonnement d'une maison d'édition*, ed. Darnton and Schlup (Neuchâtel: Editions Gilles Attinger, 2005), 375-404.

[8] For preliminary reflections on the STN's role in cultural transfers see Mark Curran, "The Société typographique de Neuchâtel and networks of trade and translation in eighteenth-century francophone Europe" in *Cultural Transfers: France and Britain in the Long Eighteenth Century*, ed. Ann Thomson, Simon Burrows, and Edmond Dziembowski, *SVEC* 2010:04 (Oxford: Voltaire Foundation, 2010), 257-67.

[9] This task will be undertaken in the two interpretative monographs stemming from the project, which will be published by Continuum in 2013. *The French Book Trade in Enlightenment Europe*, Volume 1 will be Mark Curran's study of the STN as a publisher, entitled *Selling Enlightenment*. Volume 2 will be Simon Burrows' study of the dissemination of enlightenment ideas and discourses, and is entitled *Enlightenment Bestsellers*.

SEAN TAKATS

# Review of The French Book Trade in Enlightenment Europe, 1769-1794: Mapping the Trade of the Société Typographique de Neuchâtel

Mapping the history of ideas against the actual reading practices of early modern Europeans has intrigued (and troubled) historians at least since Roger Chartier focused our attention on the need to consider the reader, not just the text. With the launch of [The French Book Trade in Enlightenment Europe, 1769-1794: Mapping the Trade of the Société Typographique de Neuchâtel \(FBTEE\)](#), project directors Simon Burrows and Mark Curran offer stunning new opportunities to probe this old problem. Although only recently publicly launched in June 2012, the site has already attracted wide renown thanks to its soft launch on H-France listserv in September 2010 and through persistent outreach like Burrows and Curran's July 2011 presentation at the Library of Congress.

FBTEE provides the first broad digital access to the archives of the Société Typographique de Neuchâtel (STN), a Swiss publisher and bookseller active in the eighteenth century. Burrows has colorfully described STN as "an eighteenth-century Amazon[.com]," but STN owes its historical significance primarily to its status as the largest and most complete extant archive of Enlightenment publishing, not because it dominated eighteenth-century book sales in the way that

Amazon does today. Because it operated beyond the reach of French censors, STN is a particularly important resource for understanding the significance and circulation of illegal works, the so-called "philosophical" books that ranged from audacious attacks on royal and religious authority to outright pornography (tartly dismissed by Rousseau as books "to be read with one hand"[1]), the two genres increasingly one and the same as the *ancien régime* collapsed. Yet because STN also dealt in uncontroversial literature, often as a counterfeit publisher, its extensive records have proven valuable to historians interested in tracing book sales and ownership where other, traditional evidence like probate inventories generally fails to shed much light.

From the traditional historian's perspective, the significance of FBTEE cannot be overstated. STN in its analog form has provided fodder for countless dissertations and books, most notably the career-making oeuvre of Robert Darnton (and indeed it's quite surprising to find his name absent from the project's advisory board, even though his *Corpus of Clandestine Literature in France, 1769-1789* (Norton, 1995) comprises a portion of FBTEE's data). That scholarship in turn has supplied the necessary backbone for still more works on Enlightenment culture and plugged gaps in the work of every scholar looking for a quick figure or bit of insight on the significance of a period text. FBTEE dramatically redraws this scholarly landscape, and the days of researchers routinely mining Darnton's and others' work for such evidence have likely ended. From the perspective of interpretation, of course, it's an entirely different story. FBTEE taken alone offers little in the way of analysis, preferring instead to allow users to draw their own conclusions and integrate its rich data into their own work. Indeed it is difficult to imagine even advanced researchers grasping the significance of FBTEE without a thorough

grounding in the history of the book that already relies on the STN archives.

FBTEE provides users with a powerful interface offering a variety of ways to approach and interrogate the STN archival data. Users can browse, search, and visualize Enlightenment books and their peregrinations. Mapping Tissot's *Essai sur les maladies des gens du monde*, for example, reveals a broad swath of purchasers far from major cities, suggesting that this decidedly urbane text appealed to readers beyond urban elites. Visitors unfamiliar with STN or digital tools will especially appreciate the project's six introductory videos with over an hour of total running time. In these charming offerings Burrows cheerfully sets about the thankless task of coaxing his technophobic colleagues into gaining a toehold on FBTEE's rich data. Burrows and Curran's thoughtful set of "designer notes" likewise offer a wonderful window into the minds of the project directors and developers, describing in detail all aspects of FBTEE such as the arduous data entry process. In design and mission, FBTEE follows in the footsteps of the acclaimed César database which has tracked eighteenth-century theatrical works and performances since the late 1990s.<sup>[2]</sup> Unlike César, however, FBTEE showcases the intervening decade's rise of open-source software (and philosophy). Its use of the D3.js maps and visualization library results in an especially polished appearance.

From a usability standpoint, the site makes certain compromises in the name of broader accessibility that may in fact limit its utility for some researchers. For example, the project directors decided to assign English-language keywords to works, which interjects an unnecessary layer of translation for much of FBTEE's target audience, who undoubtedly will think first in terms of the French originals, regardless of mother tongue. Fortunately Burrows and Curran also offer the

ability to search by "Parisian category." The site sports an interface incongruously reminiscent of desktop software in a concession perhaps aimed at attracting an audience unfamiliar with web-based tools. The pervasive use of mile-long drop-downs prohibits quick searches for known works. Predictive text inputs would provide the same effect, but less intrusively. It's not readily apparent what the FBTEE team used for its bibliographic metadata model, and its book records are lamentably only exportable via Thomson Reuters's narrow and brittle RIS format. Finally, there's no application programming interface (API) to allow third parties easily to build upon Burrows and Curran's hard work.

I am eager to add, however, that all of these quibbles are more than compensated by Burrows and Curran's generous decision to offer the entire FBTEE dataset for download, albeit subject to an ominously lengthy end user license agreement (EULA). Fortunately the gist of the EULA is that the University of Leeds, the site's hosting institution, grants free use of the FBTEE data for teaching, research, and other non-profit activities, and that the project directors expect that the site (and one of Darnton's contributing works) be credited in any resulting scholarship. This act of intellectual generosity is all the more unusual coming from humanists, who overwhelmingly view research data as wholly proprietary.

The FBTEE project isn't yet finished, in the sense that Burrows and Curran have ambitious plans for expanding its scope and reach. They write that they're interested in adding "a web-based interface through which students and scholars will be able to input their own research data and findings on the production, dissemination, or reception of print in eighteenth-century Europe and the Americas."<sup>[3]</sup> They might also consider the integration of a collaborative bibliography of scholarship related to STN. In addition to underscoring the

significance of FBTEE, such a bibliography would also highlight the important new research — like Jeffrey Freedman's *Books Without Borders in Enlightenment Europe: French Cosmopolitanism and German Literary Markets* (University of Pennsylvania, 2012) and Curran's own *Atheism, Religion and Enlightenment in Pre-Revolutionary Europe* (Boydell and Brewer, 2012) — already made possible by this pathbreaking project.

**Notes:**

[1] Robert Darnton, *The Forbidden Best-Sellers of Pre-Revolutionary France* (New York: W. W. Norton & Company, 1996), 222.

[2] *Calendrier électronique des spectacles sous l'ancien régime est sous la révolution*, <http://www.cesar.org.uk>

[3] “A Platform for Teaching and Further Research”, n.d., <http://chop.leeds.ac.uk/stn/database/platform.html>

# Mapping Texts

## PROJECT OVERVIEW

### **Mapping Texts: Visualizing American Historical Newspapers**

Andrew J. Torget & Jon Christensen . . . . . 70

## CASE STUDY

### **Building New Windows into Digitized Newspapers**

Andrew J. Torget & Jon Christensen . . . . . 73

## REVIEW

### **Review of Mapping Texts**

Robert Nelson . . . . . 81

# Mapping Texts: Visualizing American Historical Newspapers

[Mapping Texts](#) is an ambitious project with a simple mission: to experiment with new methods for finding and analyzing meaningful patterns embedded within massive collections of digitized historical newspapers.

Why do we think this is important? Because, quite simply, historical newspapers are being digitized at a rate that is rapidly overwhelming our traditional methods of research. The [Chronicling America](#) project, for example, recently digitized its 5 millionth newspaper page, and predicts that more than 20 million pages will be available within a few years. Numerous other programs are also digitizing newspapers at a rapid pace worldwide, making hundreds of millions of words from the historical record readily available in electronic archives that are reaching staggering proportions.

Such enormous collections offer tantalizing new possibilities for humanities research. Yet without tools and methods capable of sifting meaningful patterns from such massive datasets, the challenges of working with digitized newspapers are becoming equally overwhelming. Researchers, for example, too often find themselves

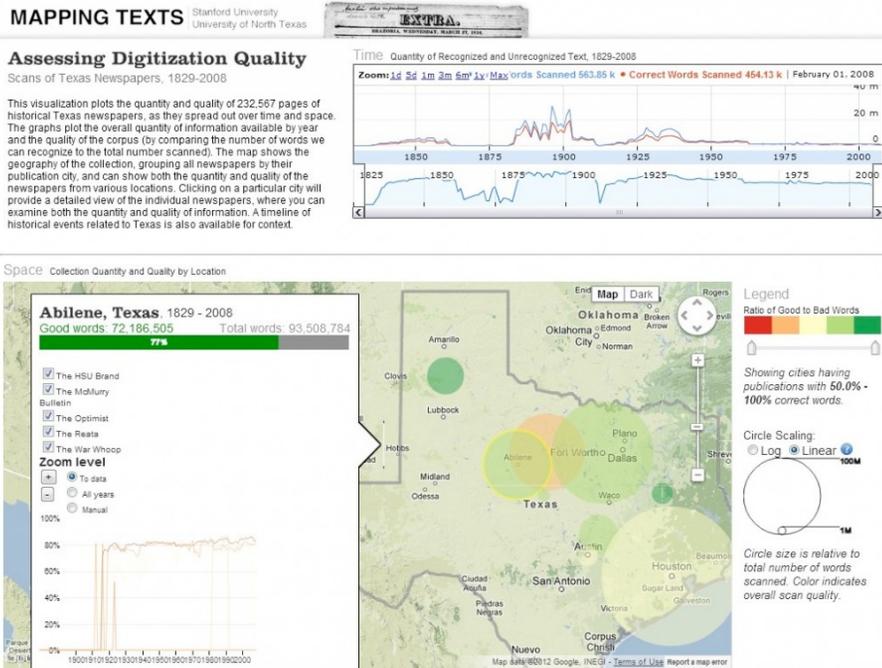
confined to exploring such archives through basic text searches (which, when they produce several million hits, offer too many results to analyze in any meaningful way by hand). And scholars invariably have no ability to evaluate basic metrics (such as how much data is available from a particular time and place, or the quality of the OCR — optical character recognition — digitization process) about a given online collection. Harnessing the promise of digitized newspapers, in other words, requires building more transparent windows into the tremendous wealth of such archives.

Our purpose with [Mapping Texts](#), then, has been to experiment with developing new methods for enabling scholars to sift, sort, and explore digitized historical newspapers for their research. To that end, we have attempted to combine the two most promising methods for analyzing large-scale datasets: data- and text-mining (for discovering meaningful patterns embedded in large bodies of text) and data visualization/mapping (for grouping, discovering, analyzing, and making sense of those patterns). Working with a collection of about 232,500 pages of digitized historical newspapers, we produced two interactive interfaces:

1. “[Mapping Newspaper Quality](#)” maps a *quantitative* survey of the newspapers, plotting both the volume and quality (OCR recognition rates) of information available in the digitized collection. Through graphs, timelines, and a regional map, users can explore these metrics for any particular time period, location, or newspaper. Clicking on individual newspaper titles also allows users to jump from “distant” to “close” readings of the texts.
2. “[Mapping Language Patterns](#)” maps a *qualitative* survey of the newspapers, plotting major language patterns embedded in the collection. For any given time period, geography, or newspaper title, users can explore the most common words (word counts), named

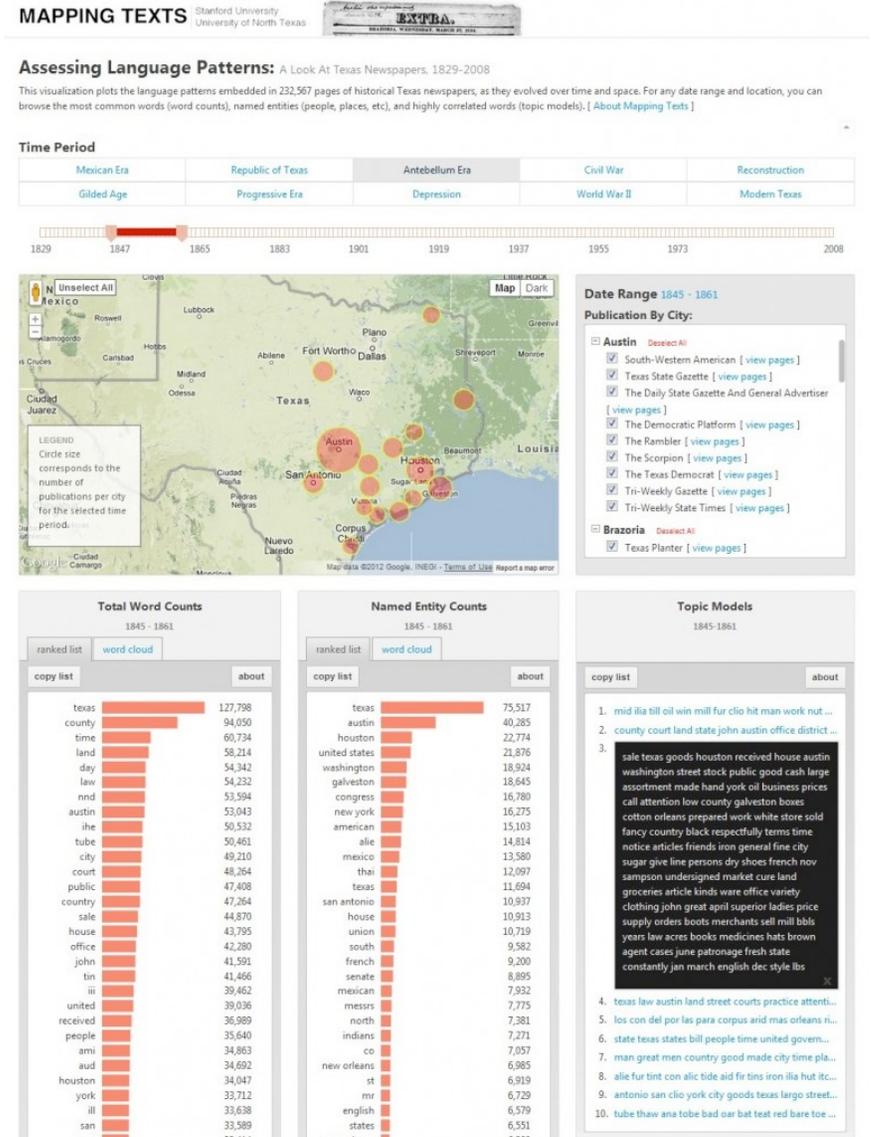
entities (people, places, organizations), and highly correlated words (topic models), which together provide a window into the major language patterns emanating from the newspapers. Clicking on individual newspaper titles also allows users to jump from “distant” to “close” readings of the text.

represented the entire corpus available to us — or any researcher — accessing UNT’s digital newspaper archive when we began the project during the fall of 2010.



### Mapping Newspaper Quality

These two interfaces are built on top of the large archive of historical newspapers digitized by the University of North Texas (UNT) as part of the [Chronicling America](#) project and UNT’s [Portal to Texas History](#). We selected this archive for a number of reasons: with nearly a quarter million pages, we could experiment with scale; the newspapers were digitized to the standard set by Chronicling America, providing a uniform sample; the Texas orientation of all the newspapers gave us a consistent geography for our visualization experiments. It also



### Mapping Language Patterns

The project was, at base, an experiment to see what we could discover about the breadth and depth of a single electronic newspaper archive, and what that might tell us about other similar archives. The project's interfaces are meant to be used in tandem, with the hope that researchers will combine insights from the two in order to better sift through these collections and perhaps discover previously hidden connections in the newspapers.

This work depended heavily on collaborations between scholars at UNT and Stanford University — UNT's Rada Mihalcea and Stanford's Geoff McGhee, in particular. Please see the [project website](#) for a full listing of the team behind the project.

# Building New Windows into Digitized Newspapers

In building [Mapping Texts](#) we wanted to create more transparent windows into the extraordinary wealth of information available in online archives of digitized historical newspapers. We wanted, for example, to be able to *see* how much information was available for any particular time and place, and then *measure* just how much of that information was still recognizable – and thus useable – after the digitization process. We also wanted to be able to see the major language patterns coming from these datasets, so that we could use “distant” readings of such massive collections as a way to determine which individual newspapers would likely yield the most useful information from “close” readings. We wanted, in other words, more finely grained methods for indexing both the quantity and quality of information in these archives as they spread out across both time and space.

These are critical considerations for anyone conducting research. If a scholar cannot judge with any accuracy whether a particular archive has a sufficient amount of information about a particular time period, region, or subject, it becomes tremendously difficult – if not impossible – to determine whether that archive’s holdings would yield useful

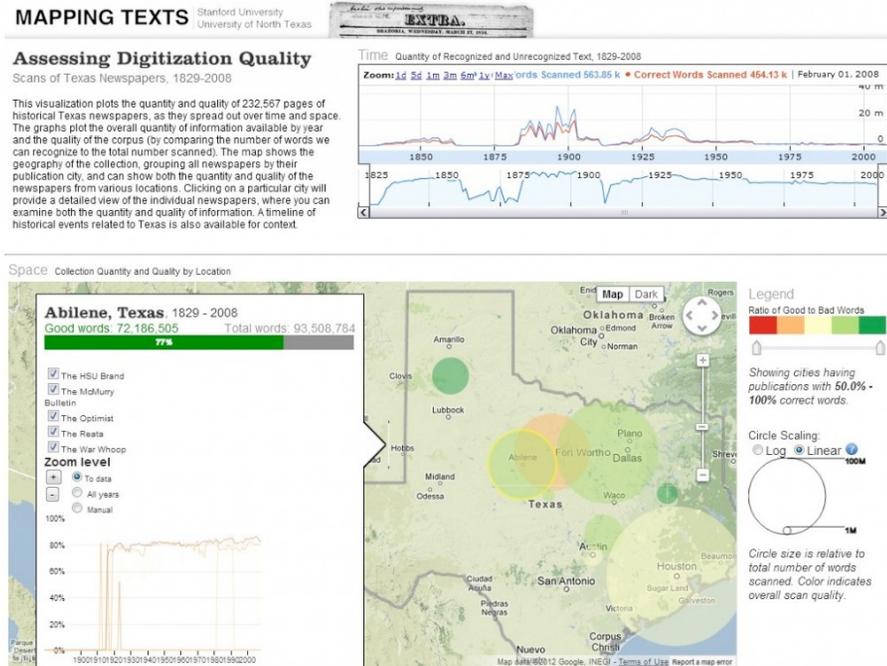
material for a given research project. And in most online archives of historical newspapers, those basic metrics are rarely accessible. If, however, we could develop new ways to expose and analyze those crucial pieces of information, we could enable scholars to make more informed choices about which datasets would be most useful to them in their research.

Our underlying purpose, then, was simply to enable better research in historical newspapers. And to that end we experimented with an online archive of 232,500 pages of historical Texas newspapers digitized by the University of North Texas’s [Portal to Texas History](#). The resulting two interfaces – “[Mapping Newspaper Quality](#)” and “[Mapping Language Patterns](#)” – each explore a different side of the archive’s holdings, one a quantitative view and the other a qualitative view. When used together, our hope is that the two interfaces will offer one possible model for how we might expose more clearly the full breadth and depth of information available in large collections of digitized historical newspapers.

## Mapping Newspaper Quality

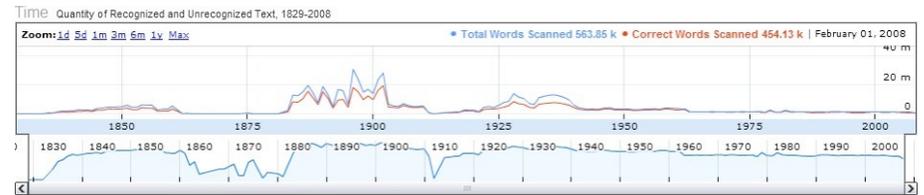
This interactive visualization plots the quantity and quality of information in the digital archive by geography and time periods, using both to survey the amount of information available for any given time, place, or individual newspaper. The entire corpus was broken down into recognizable words (words that could be matched against dictionary entries) and “noise” (words that were garbled by the OCR digitization process), and then plotted by time and location at the macro-level (that is, Texas as a region), the local level (individual cities), and the micro-level (individual newspaper titles). By allowing researchers to expose and explore both the volume and reliability of information available to them in a database of digital newspapers, our

hope has been to enable scholars to make more informed choices about what research questions they can and cannot answer from a given collection of historical newspapers. (For more detail about the OCR process, the problem of “noise,” and our methods in recognizing and counting words, please see our [White Paper](#).)



### Mapping Newspaper Quality

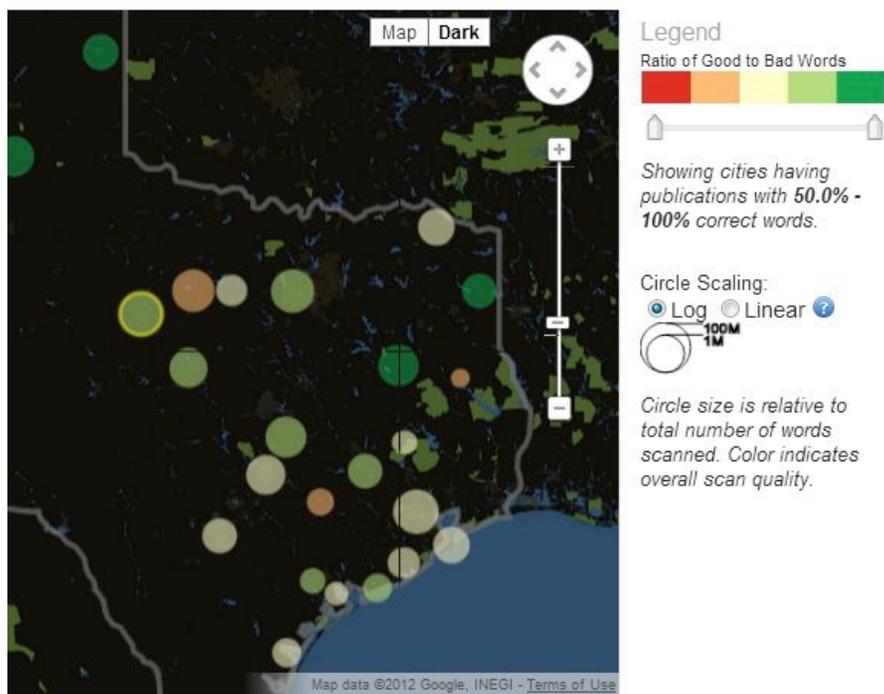
Even a casual look at [“Mapping Newspaper Quality”](#) reveals important patterns. Although the entire newspaper archive represents 1829 to 2008, visualizing the quantity of data on a timeline reveals that newspapers within two smaller eras – 1883 to 1911 and 1925 to 1942 – vastly outnumber the representation from any other period (there is also a slight rise for the 1840-1860 era).



### Mapping Quality Timeline

This reveals an important, although easily overlooked, fact about most digital newspaper archives: they tend to be strongest in subjects and time periods that garner grant funding. In this particular case, the UNT libraries digitized the bulk of their newspaper content according to the requirements of [Chronicling America’s](#) 2007 digitization grants (which concentrated on 1880-1910) and a private foundation grant (focused on the early twentieth century). Visualizing the volume of data available by time periods thus suggests that scholars of the Gilded Age and Progressive Era would be far better served by this particular dataset than would scholars of other time periods.

Adjusting the timeline also affects the other major window into the content: an interactive map of Texas. For the visualization, all the newspapers in the database were connected by their publication city, so the map shows the geographic distribution of the newspaper content by city. This can be adjusted to show the varying levels of quality in the newspaper corpus (by adjusting the ratio bar for “good” to “bad” words) in order to find the areas that had higher or lower concentrations of quality text. The size of the circle for each city shows their total content relative to one another – which the user can switch from logarithmic view (the default view, which gives a better sense of the proportional difference between locations) to a linear view (which provides a better sense of the absolute difference between locations).

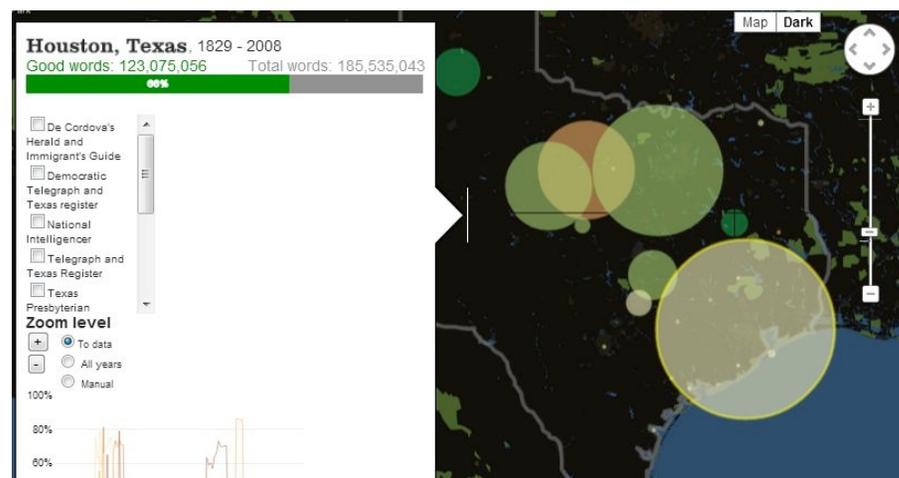


Distribution of Information by City

Viewing the database geographically reveals that two locations dominate the collection: newspapers from Houston and Ft. Worth. Combined, these two locations outstrip the volume of information available from any other location in Texas, which is interesting in part because neither of those locations became dominant population centers in Texas until the post-World War II era (and therefore after the 1883-1911 and 1925-1940 time periods that compose the majority of the newspaper content). This would suggest that the newspapers of rural communities, where the majority of Texans lived during the Gilded Age and Progressive Era, are underrepresented among the newspapers in this collection, and that urban newspapers – and therefore urban concerns – are likely overrepresented. While scholars of urbanization would be well served by the collection, scholars

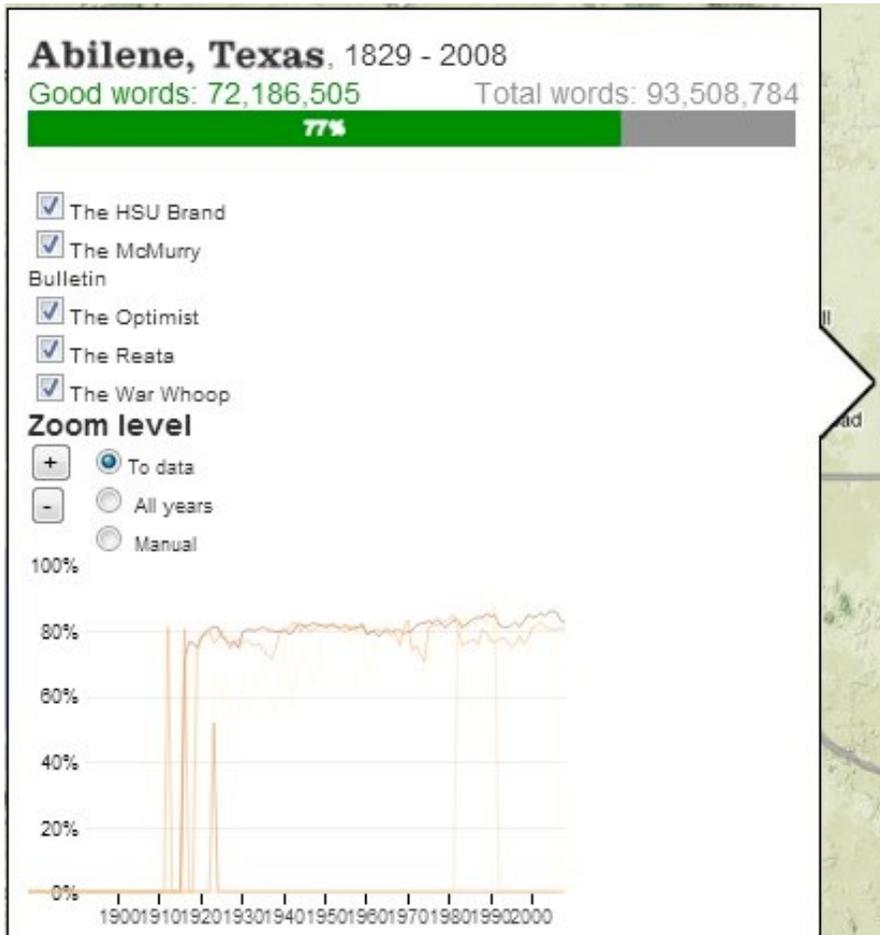
interested in rural developments would be advised to be wary of this imbalance when conducting research in the archive.

The third major window into the collection is the detail box that, for any given location (such as Abilene, Texas), provides a bar measuring the ratio of recognized words against noise, a complete listing of all the newspapers that correspond to that particular location, and the OCR recognition rates for individual newspapers:



Volume of information from Houston and Ft. Worth compared to other Texas cities.

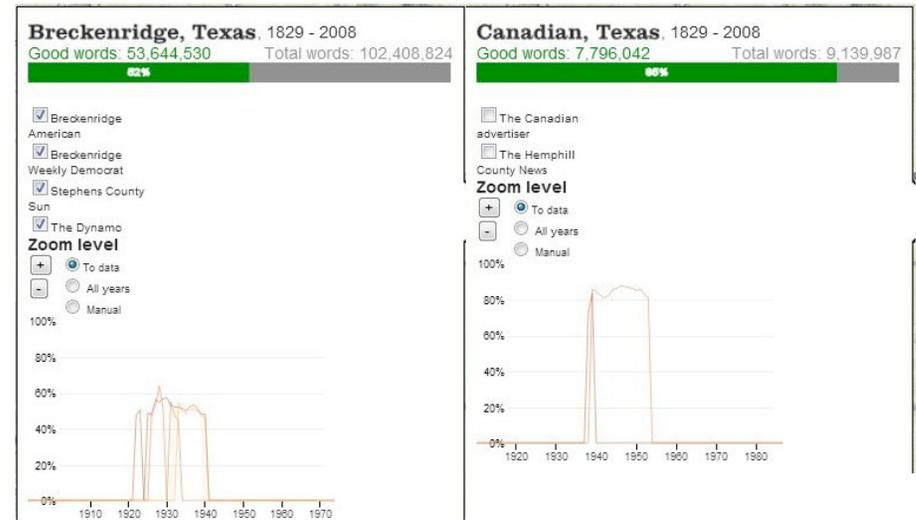
The detail box [next page] also provides access to the original newspapers themselves (as clicking on any given newspaper title will take the user to the original pages on the UNT's [Portal to Texas History](#)), allowing users to jump from “distant” to “close” readings.



Detail window of information available from Abilene, Texas.

Exploring the various geographic locations with the detail box reveals more useful patterns about the information available in the dataset. Although Houston and Ft. Worth represent the locations with the largest quantity of available data, they are not the locations with the highest quality of available data. The overall recognition rate for the OCR of Houston newspapers was only 66 percent (although this varied widely between various newspapers), and for Ft. Worth the overall rate was 72 percent. By contrast, the newspaper in Palestine, Texas,

achieved an 86 percent quality rate, while the two newspapers in Canadian, Texas, achieved an 85 percent quality rate. At the lowest end of quality was the OCR for newspapers from Breckenridge, Texas, which achieved only a 52 percent rate. Scholars interested in researching places like Breckenridge or Houston (which were both highly involved in the expansion of the Texas oil industry during the early twentieth century), then, would need to consider that anywhere between a third to fully half of the words from those newspapers were rendered unrecognizable by the OCR process. Scholars who decided to focus on newspapers from Palestine or Canadian, on the other hand, could rely on the high quality of the digitization process for their available content.

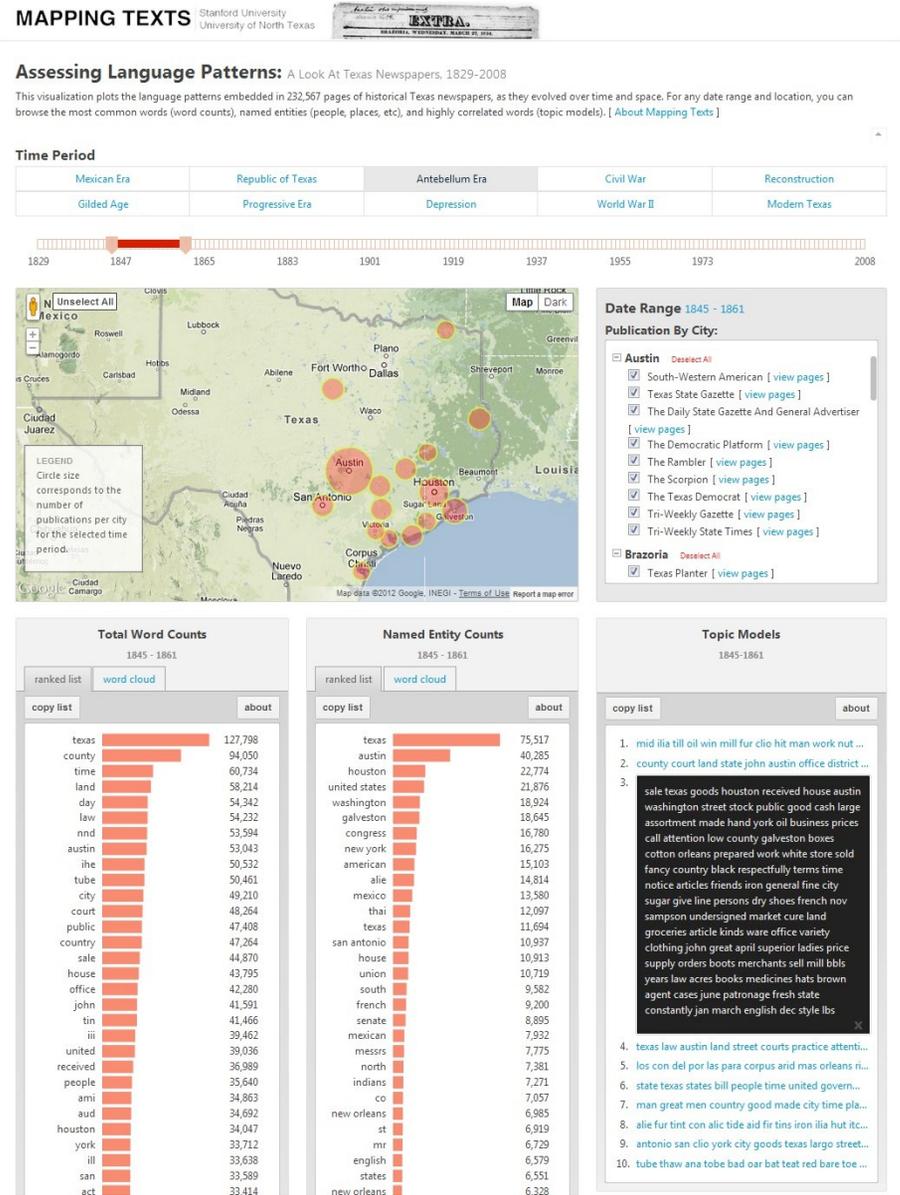


Detail windows on data available from Breckenridge and Canadian, Texas.

# Mapping Language Patterns

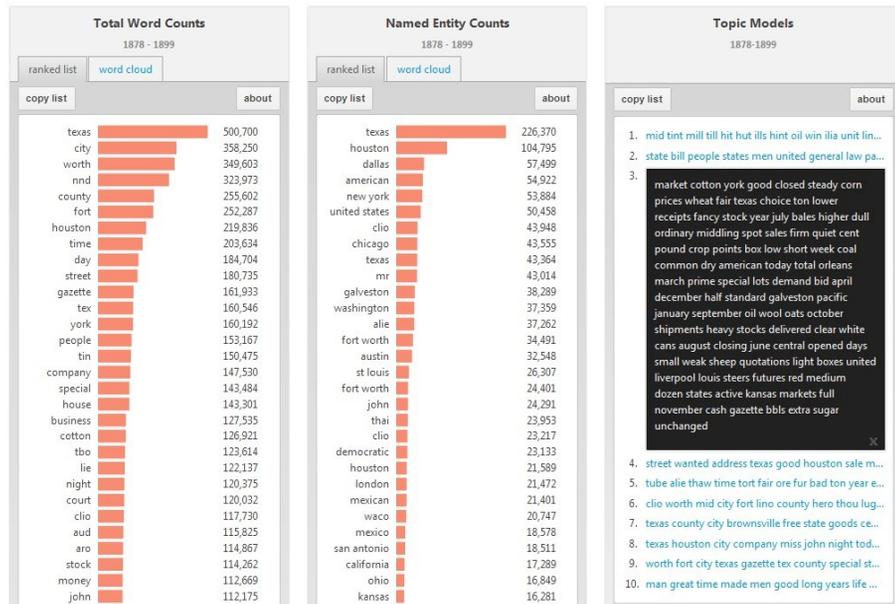
For our second interface, “[Mapping Language Patterns](#),” we built a qualitative survey of the language patterns in our digitized newspaper collection. We chose to focus on three of the metrics most widely used by humanities scholars for surveying language patterns in large bodies of text: word counts (ranked counts of all the words, minus extremely common “stop words” such as articles and conjunctions), named entity counts (ranked counts of named entities, such as people, places, and organizations), and topic models. (For greater detail about these metrics – particularly topic modeling – and our methods in generating them, see our [White Paper](#).)

Just as with our quantitative model, the user can select any time period from 1829 through 2008. For several reasons, we have also included pre-set buttons for historically significant eras in Texas and U. S.-Mexican borderlands history (“Mexican Era,” “Republic of Texas,” “Antebellum Era,” “Civil War,” and so on) which, if clicked, will automatically reset the beginning and end points on the time slider to those particular eras. (One of the key reasons for pre-setting the historical eras was that, due to the limits of current technologies, all topic models had to be preprocessed for specific time frames.) Once the user has selected a time frame, they can also customize the geography they want to examine. Based on the timeline selection, the map populates so that the user sees all the cities that have publications from the time period they selected. The user, then, can choose to examine all the language patterns emanating from all the newspapers relevant to their time period, or they could customize their selection to particular locations (such as Houston) or even a single newspaper title.



Mapping Language Patterns

Once a user has selected a time frame and geography, they can then examine the three major language patterns which are displayed on the map in their own “widgets”:



### Language Pattern Widgets

In the word counts and named entity counts widgets, there are two ways to explore the language data: (1) as a ranked list – with the most frequently appearing words at the top followed by a descending list – which reveals the most frequently used terms in the collection, and (2) as a word cloud that shows the constellation of words being used, with the size of the font indicating the relative rank of words in terms of frequency. Word clouds have become some of the most common and popular methods of displaying word counts, and we see a great deal of value in its ability to contextualize these language patterns. But we have also found that ranked lists of these same words are highly effective and offer a more transparent, direct method for displaying how these words relate to one another in terms of frequency.

In the topic model widget, the user is offered the top ten most relevant “topics” associated with a particular date range. Within each topic is a list of 100 words that have a statistical relationship to one another in the collection, with the first word listed being the most relevant, the second being the second-most relevant, and so on. The 100 words are truncated for display purposes, but clicking on any given topic will expand the word list to encompass the full collection, which allows the user to parse and explore the full set of topic models.

Each topic’s collection of words is meant to expose a theme of sorts that runs through the words in the newspapers selected by the user. Sometimes the topic is a collection of nonsensical words (like “anu, ior, ethe, ahd, uui, auu, tfie” and so on), when the algorithm found a common thread among the “noise” (words jumbled by the digitization process) and recognized a commonality between these non-words, which it then grouped into a “topic.”

More often, however, the topic models group words that have a clear relationship to one another. If, for example, the user were to select all the newspapers from the Republic of Texas era, one of the topic models offered includes "Texas, government, country, states, united, people, mexico, great, war . . ." which seems to suggest that a highly relevant theme in the newspapers during this era were the international disputes between the United States and Mexico over the future of the Texas region (and the threat of war that came with that). That comports well, in fact, with what historians know about the era. What is even more revealing, however, is that most of the other topic models suggest that this was only one – and perhaps even a lesser – concern than other issues within the newspapers of 1830s and 1840s Texas, such as matters of the local economy (“sale, cotton, Houston, received, boxes, Galveston”), local government (“county, court, land, notice, persons, estate”), and social concerns (“man, time, men, great, life”),

which have not received nearly as much attention from historians as the political disputes between the United States and Mexico during this period.

**Topic Models**  
1836-1845

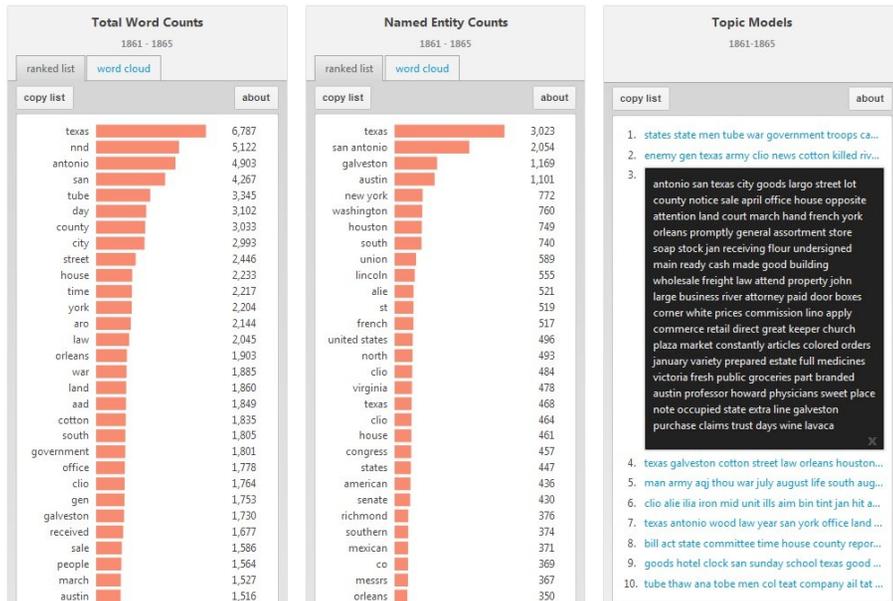
copy list about

1. ana arid anu ior ethe ahd uui auu tfie uiu iuc sss t...
2. man time men great life made good long water f...
3. cent cats block adv texas lot supply lbs kinds scar...
4. tube alie ilia fur con aid tint fir thaw fit win lion li...
5. house bill act senate committee sec time county ...
6. county court land notice persons estate john law ...
7. sale cotton houston received boxes galveston co...
8. paper state public free country dollars year work ...
9. texas government country states united people  
mexico great war made president citizens  
present state mexican congress general men  
gen time public power subject santa part army  
place annexation nation union means received  
republic question treaty make good american  
arrived national anna troops character opinion  
policy col letter houston washington indians  
city order san interest important man party  
executive officers large force measure foreign  
purpose years territory give number laws  
constitution peace found honor independence  
duty enemy hope portion west true days  
orleans doubt editor fact nations regard  
amount british object england cotton left law  
trade rights friends political western favor
10. city texas houston law office sale john notice cou...

Topic models for Republic of Texas Era (1836-1845)

Some of the patterns fit neatly into standard interpretations of these eras. Discussions of war, for example, dominate the topic models for the Civil War, as do concerns about the collapsing cotton market during the Gilded Age. And across the different eras, concerns about public issues – such as economics, politics, and legal matters – dominate the conversations in the newspapers and can be seen most readily in the topic models.

More remarkable are the anomalies that emerge which do not fit neatly into how historians have typically described these eras. In comparing the ranked lists of named entities over time, for instance, there appears to be a focus among Texas newspapers on “New York” that began to rise during the 1840s and peaked during the late nineteenth and early twentieth centuries (during which time “New York” ranked consistently among the most frequently named locations). This may be a reflection of how intertwined the Texas economy of those years was with markets and financiers in New York; it may reflect the dominance of New York advertisements in the newspapers; or it may be related to other causes altogether. During the Civil War, “San Antonio” emerged as a highly prominent phrase in word counts, named entities, and topic models – which is remarkable, since the city played no prominent role in the state’s war effort. And it was not until the Depression of the 1930s that sports becomes a significant theme in the Texas newspapers of this archive (which may reflect the fact that college newspapers make up a large proportion of the archive during those particular years), even though sports had been an important part of early twentieth century life in the state before the collapse of the American economy. Each of these patterns could suggest possible lines of new research and study for humanities scholars and help guide their forays into the digital archive.



Topic Models for Civil War Era (1861-1865), where "San Antonio" appears prominently.

Digital humanities scholars and technologists are rapidly developing new tools for exploring, analyzing, and interpreting the massive new databases of text, and “Mapping Texts” does not stand alone. It was designed to fit into an emerging ecosystem of tools, processes, and methods in text-mining, analysis, and visualization. The true proof of the utility of this experiment, however, will come if these tools prove useful in shaping the questions that scholars ask, in guiding their use and interpretation of these digital archives, and if any of the array of analytical and visualization technologies that we have combined here are more widely adopted for exploring the vast digital archives that are increasingly becoming available for humanities scholars.

# Review of Mapping Texts

[Mapping Texts](#) addresses a pressing problem and an exciting opportunity that many working in digital humanities are tackling: now that we have massive amounts of historical evidence digitized, what new questions can we pose and what new knowledge can we produce? Developed through a collaboration of two teams at the University of North Texas and Stanford University — the former led by Andrew Torget and including Mark Phillips, Rada Mihalcea, and Eliza Tze-I Yang; the latter led by Jon Christensen and including Cameron Blevins, Rio Akasaka, Geoff McGhee, Yinfeng Qin, and Jason Ninxuan Wang — the project seeks, to quote their [white paper](#), “to help scholars develop new tools for coping effectively with the growing challenge of doing research in the age of abundance” by marrying two common methods: text mining and geovisualization.

As the creators of Mapping Texts note, these two methods are typically used independently of one another to uncover patterns in and make sense of massive amounts of digitized evidence. In Mapping Texts these methods are combined in two provocative visualization tools that aim to explore a very large corpus: nearly a quarter of a million pages

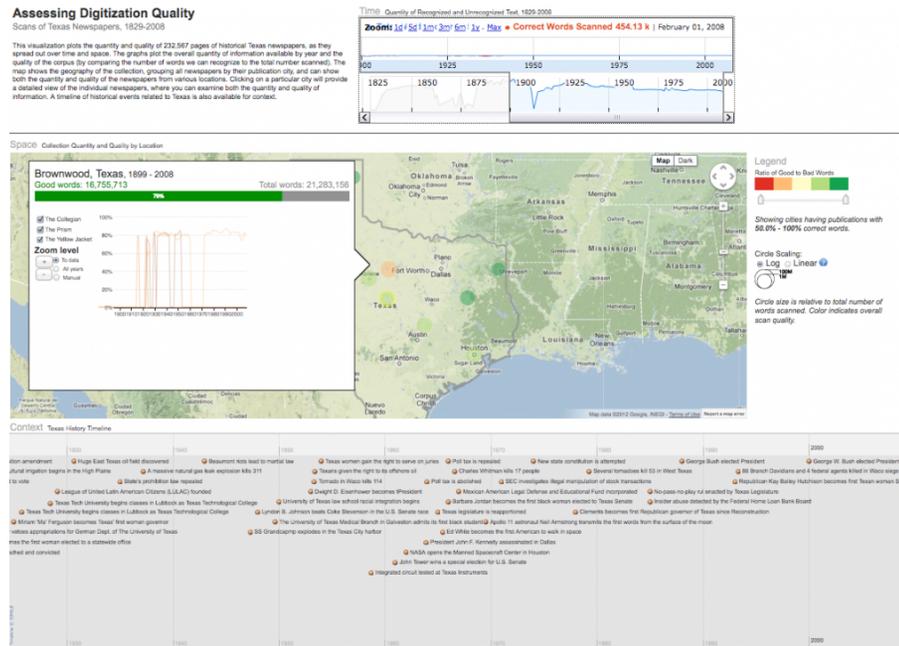
from 100 Texas newspapers from the [Chronicling America](#) project dating from 1829 to 2008.

The first of these visualizations, “[Assessing Digitization Quality](#),” measures the quantity and, more importantly, assesses the quality of OCR for these newspapers. It presents three visualizations. The first is a line chart that shows the overall quantity of available newspaper text (measured in number of words) and the overall quality of that OCRred text (measured as the ratio of recognizable words to total words) over the nearly two centuries these newspapers collectively span. The second is a map of Texas visualizing these quantities over space. The third is a handy timeline that provides some basic historical context by showing major events in Texas history.

Together these visualizations are certainly helpful for getting a quick sense of when and where the newspapers holdings for the Texas portion of Chronicling America are concentrated. They are also very useful for gauging the accuracy of the digitized text for particular time periods, locales, and newspapers. The assessments the developers of Mapping Texts have done reveal the accuracy of the OCR to vary substantially not just across time — with newspapers from the latter half of the twentieth century having significantly greater accuracy — but also across space. For instance, newspapers from the small towns of Palestine and Canadian have, respectively, OCR accuracy of 85% and 86%; on the other end of the spectrum, the newspapers of the town of Breckenridge only have an accuracy rate of 52%.

Beyond assessing the OCR quality of the newspapers, the creators of Mapping Texts applied a number of techniques to algorithmically correct errors. In this they have done a great service. While their techniques could only correct about a fifth of the identified errors, given the enormous number of errors that exist that number of corrections is substantial, certainly many millions of words. One hopes

that these techniques will be further assessed, refined, and eventually applied to the entirety of Chronicling America to improve the overall quality of that valuable archive.



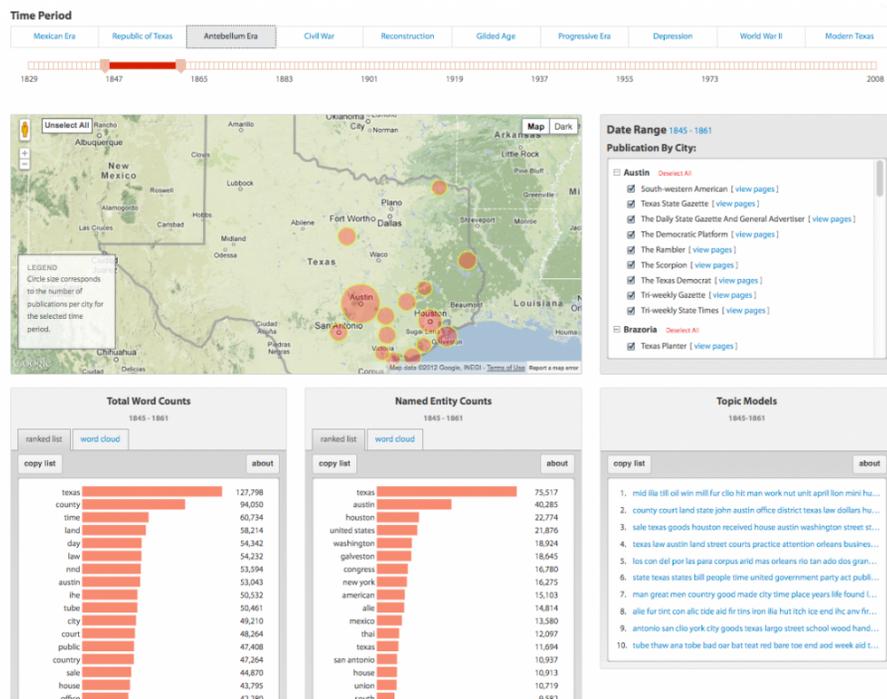
## Assessing Language Quality Provides Measurements of Quantity and OCR Quality for Locations and Time Periods

Measures of both the quantity and quality of such massive digitized corpora should be taken into account, the creators of Mapping Texts suggest, by anyone hoping to apply text-mining methods for analysis. Proposing in their white paper that “almost all research questions would first require a quantitative survey of the available data,” they provide a hypothetical example: “If poor imaging — and therefore OCR results — meant that ‘Lincoln’ was often rendered as ‘Linco1n’ [with a numeric 1 rather than an l] in a data set, that *should* affect how a scholar researching newspapers patterns

surrounding Abraham Lincoln would go about his or her work.” This intuitively makes sense, and “Assessing Digitization Quality” is an extremely flexible and quick way to gauge the OCR quality of these newspapers at particular moments and places. But in what practical ways should a researcher compensate for these varying levels of quality? For example, what should a researcher interested in applying text-mining and visualization techniques to the newspapers of Breckenridge with its low 52% accuracy do differently than a researcher interested in Palestine or Canadian with their percentages of accuracy in the mid 80s? What difference does that 33% make? Can a given text-mining or visualization technique be successfully applied to the latter towns and not to the former one? Mapping Texts raises but does not address these important and admittedly complicated and challenging questions, leaving it to researchers to determine how they might take these measures of OCR quality into account in their research.

These measures aren’t taken into account by the creators of Mapping Texts themselves in the text-mining visualizations they provide in the second of the site’s tools, “[Assessing Language Patterns,](#)” which aims to allow researchers to explore some of the dominant language patterns in the newspapers. It presents three word visualizations along with a slider and map to limit results to particular spans of time and locations. The first provides a tally of the 50 most frequently used words for a selected time period, place, and newspaper (or some combination of places and newspapers). The second is much the same but limits the words to algorithmically identified “named entities” (persons, places, organizations, etc.). The third presents ten “topics” from a substantial number of topic models that were produced for preselected time periods (the Mexican Era, the Republic of Texas, the antebellum period, etc.) from all newspapers and from the newspapers of each city.[1] These views can provide a suggestive if

somewhat rudimentary sense of some of the dominant language patterns across time and space. For instance, for each era “Washington” appears in the named entity list but its relative prominence changes substantially over time, from the 14th most frequently used named entity during the Mexican Era, growing in importance through the antebellum decades (5th) and Civil War (6th), and declining over the next century and a half to a low of 37th during the postwar era.



### Assessing Language Patterns Provides Three Different Language Metrics

While certainly interesting, the kinds of patterns that can be detected using these views must be assessed very critically given both the techniques that are being employed and the types of newspapers that

are present in *Chronicling America* for the different eras. For instance, the relatively low ranking of “Washington” during the postwar period likely conveys less about its importance during those decades than about the composition of the corpus: six of the ten newspapers for that period are college student newspapers, hardly representative of modern newspapers in Texas or straightforwardly comparable to the types of newspapers digitized for the out-of-copyright era. The counts of unigrams provided by Mapping Texts are undoubtedly useful for some researchers and some research questions, but they are a very limited metric for textual analysis; longer phrases are typically much more revealing as are more sophisticated text-mining techniques.

Another limitation of “Assessing Language Patterns” is the inability to move from any of the provided views into the newspapers themselves to study a suggestive or intriguing pattern in greater depth, in other words to move from the distant readings it affords to close readings of the texts from which it’s derived. For instance, providing a straightforward route from the topics into the texts themselves would have been welcome — functionality akin to that provided by [David Blei’s topic model of the American Political Science Review](#) or in my [“Mining the Dispatch.”](#) While the list of most frequent words for topics in a topic model often can provide a clear sense of a particular issue, theme, or genre, they are just as often ambiguous or even opaque taken by themselves. Sometimes reading them without reference to representative documents can be less like sound research and more like a Rorschach test — when read in isolation what one sees might reveal more about one’s own preoccupations and knowledge than about the documents themselves. It is always useful and sometimes essential to be able to move from the word distribution of a topic to the documents that are highly representative of that topic to get a more subtle sense of its substance and meaning, a point Lisa Rhody thoughtfully illustrates in [a recent blog post](#). (That said, it’s unclear

how the creators of Mapping Texts defined a document. I don't believe that the newspapers in *Chronicling America* are segmented by article, so they may have used pages or individual issues as the "documents" for their models. In that case it would be understandable if still regrettable not to provide a way of moving from topics into the texts.)

The creators of Mapping Texts should be applauded for taking on the daunting task of grappling with the entirety of an archive of enormous scale and scope and of a deeply imperfect nature. If Mapping Texts falls short in some respects, it still contributes to the challenging and slow project of developing new techniques to analyze the massive amount of digitized evidence that is now available. In particular, it does a service by highlighting the significant challenge of applying text-mining techniques to documents rife with OCR errors.

**Notes:**

[1] Scott Weingart recently provided [a very useful and thoughtful overview](#) of topic modeling in digital humanities for those unfamiliar with this technique.

---

# Contributors

**Matthew Morse Booker** is associate professor of American and environmental history at North Carolina State University. His work examines the intersection between human beings and the natural world in North America, with a particular focus on the coastal regions. His first book, *Between the Tides: Layers of History in San Francisco Bay* (University of California Press, 2013) investigates ownership and ecology in a contested urban and natural space. Matthew's current research traces the history of food production in American industrial cities of the late nineteenth and early twentieth centuries. From 2008-2010 Matthew was a founding member of a collaborative student-faculty experiment at Stanford University, the Spatial History Project, using mapping tools to creatively frame and visualize historical problems. He remains an affiliate scholar. Matthew is also active in the Research Triangle Digital Humanities community.

**Simon Burrows** is currently Professor of Modern European History at the University of Leeds and instigator and principal investigator of the AHRC-funded French Book Trade in Enlightenment Europe (FBTEE) project. He is author or co-editor of six books and numerous articles, most of which concentrate on print culture in the period 1750-1850. In January 2013 he will be taking up a Professorship in History at the University of Western Sydney in Australia.

**Jon Christensen** is a founder and principal investigator in the [Spatial History Project](#) and the [City Nature](#) digital humanities project at Stanford University, and former executive director of the [Bill Lane Center for the American West](#). In the fall of 2012, he will join the [History Department](#) and the [Institute of the Environment and Sustainability](#) at the University of California, Los Angeles, where he looks forward to working with the [Center for Digital Humanities](#) as well.

**Mark Curran** is currently Munby Fellow in Bibliography at the University of Cambridge and author of *Atheism and Christianity in Enlightenment Europe* (Boydell, 2012). In October 2012 he will take up a three year lectureship at Queen Mary University of London. From 2007-2011 he was Research Fellow on the French Book Trade in Enlightenment Europe project, which he has been involved in shaping since the pilot project in 2006. He is co-author with Simon Burrows of the FBTEE database.

**Stuart Dunn** is a Lecturer in the Department of Digital Humanities at King's College London, where he was previously a Research Associate and Research Fellow. He teaches digital preservation, cultural heritage, and archaeology and material culture. Stuart graduated from the University of Durham with a PhD in Aegean Bronze Age Archaeology

in 2002, conducting fieldwork and research visits to Melos, Crete, and Santorini. His interests include ancient geography, the representation and enforcement of boundaries and frontiers, and the representation of human motion in 3D digital reconstruction of cultural heritage, and in museums. Most recently he has worked on the virtual reconstruction of social space in Southern British Iron Age round houses, and on the geography and digital representation of historic English place names. He is also interested in the theory and practice of digital communities using social media, and the development of digital infrastructures to support collaborative research.

**Karl Grossner** is a Digital Humanities Research Developer at Stanford University Libraries. He earned a Ph.D. in Geography at University of California, Santa Barbara in 2010. Karl's principle research interests are in the representation of historical knowledge in computational models of place, viewed as the setting for activity, events, and processes. This gets him into developing ontology-driven spatial databases – geo-historical information systems that underlie digital atlases – and more generally, 'interactive scholarly works' that present historical argument. It also entails semantic computing with natural language processing methods such as topic models and semantic similarity measures. At Stanford, Karl has co-developed the ORBIS web site (<http://orbis.stanford.edu>) and is currently working on two completely unrelated projects, 'CityNature' and 'Kindred Britain,' both with early 2013 expected release dates. In the "before time" prior to entering academia he had occasionally overlapping professional careers in web software project design, network systems consulting, architectural computing, and photography.

**Elijah Meeks** is the Digital Humanities Specialist at Stanford University, where he helps 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.

**Craig Mod** is a writer, designer, and publisher. He is a 2011 TechFellow award recipient as well as a 2011/2012 MacDowell writing fellow. His writing has appeared on [nytimes.com](http://nytimes.com), *New Scientist magazine*, *Codex Journal*, and other publications. He is the co-author of *Art Space Tokyo*. His current work revolves around the future of publishing and reading and how these changes are impacting education. Formerly he was a product designer at Flipboard with a principal focus on the Flipboard for iPhone product.

**Robert Nelson** is Director of the Digital Scholarship Lab at the University of Richmond and an American Studies scholar whose work focuses on the nineteenth-century United States.

**Sean Takats** is Assistant Professor of History at George Mason University and Director of Research Projects at the Roy Rosenzweig Center for History and New Media. His research focuses on early modern France, the Enlightenment, and the digital humanities. At the Roy Rosenzweig Center for History and New Media, Takats has directed Zotero, the popular research software platform, since 2006. He has also led other projects on text mining and the history of science. Takats is also currently co-director of The Encyclopedia of Diderot and D'Alembert Collaborative Translation Project, hosted at the University

of Michigan. Takats is author of *The Expert Cook in Enlightenment France* (Johns Hopkins University Press, 2011). His current research extends his interest in occupational expertise to the colonial world, where he explores the practices of collecting and synthesizing a wide range of exotic knowledge, ranging from botany to commerce to medicine.

**Melissa Terras** is the Co-Director of University College London (UCL) Centre for Digital Humanities, and Reader in Electronic Communication in the Department of Information Studies at UCL. She teaches Internet Technologies, Digital Resources in the Humanities, and Web Publishing, and supervises a range of Masters and PhD students. Her research interests involve applying computational technologies to Humanities problems, to allow research that would otherwise be impossible. As such, she's interested in – and been involved in – a variety of research areas that span many areas of Digital Humanities. Current projects include QRator, Digital Transformations, and Textal. Previous projects include Log Analysis of Internet Resources in the Arts and Humanities, Virtual Environments for Research in Archaeology, eScience and Ancient Documents, and Researching eScience Analysis of Census Holdings. Terras is the general editor of *Digital Humanities Quarterly*, and is currently Secretary of the Association of Literary and Linguistic Computing and serves as an executive of the Alliance of Digital Humanities Organizations.

**Andrew J. Torget** is a historian of nineteenth-century North America at the University of North Texas, where he directs the Digital History Lab. The founder and director of numerous digital humanities projects – including [Mapping Texts](#), [Texas Slavery Project](#), [Voting America](#), and the [History Engine](#) – Andrew served as co-editor of the

[Valley of the Shadow](#) project, and as the founding director of the [Digital Scholarship Lab](#) at the University of Richmond. The co-editor of several books on the American Civil War, Andrew has been a featured speaker on the digital humanities at Harvard, Stanford, Rice, and the National Archives in Washington, D. C. In 2011, he was named the inaugural David J. Weber Research Fellow at the [Clements Center for Southwest Studies](#) at Southern Methodist University.

**Sarah Werner** is the Undergraduate Program Director at the Folger Shakespeare Library, where she teaches courses on early modern book history, and Editor of [The Collation](#), the Folger's research blog. She is the author of *Shakespeare and Feminist Performance* (Routledge, 2001), the editor of the collection *New Directions in Renaissance Drama and Performance Studies* (Palgrave 2010), and is Textual Editor of *Taming of the Shrew* for the forthcoming 3rd edition of the Norton Shakespeare. She is currently exploring the intersections of book studies and digital tools and writing about the value of bringing undergraduates into special collections.