Digital Research Methods: Databases, Ethics, Enactments, Histories, and Processes

Megan McIntyre, Dartmouth College
Rik Hunter, University of Tennessee-Chattanooga
Kerry Banazek, University of Pittsburgh
Kelsey Cameron, University of Pittsburgh

How do ever-shifting digital methods and research sites alter the character of writing research? How might we understand the historical, practical, and ethical dimensions of such research? Taken together, these presentations offer a set of questions and provocations (and a few potential—and preliminary—answers) meant to open conversations about the specific challenges of digital writing research.

Last fall, when Rik and Megan began talking about putting this panel together, they were both fresh from the Dartmouth Summer Seminar for Composition Research. The seminar, then in its fourth year, brings together writing studies scholars from across the world to learn more about empirical, qualitative, and quantitative research methods and to develop their own projects in concert with leading researchers in the field. During the seminar, Rik and Megan talked a number of times about the difficulty of method, especially digital research methods. They also talked a bit about research concerns beyond methodology, including the very real and pressing issues of tenure, time, and failure.

Too often in conference presentations and published papers, we shy away from discussing the specific and banal concerns that present themselves when we conduct writing research. And frequently, we are afraid to admit that we are lost or that we have employed particular methods that now seem less than ideal. We offer the following pieces to open up those conversations in an effort to find ways to better represent how research is actually conducted in writing studies.

We also offer the following pieces as a gesture toward better understanding how digital writing spaces and methods may have changed our research practices. Twelve years have passed since Kathleen Yancey called on the attendees of CCCC to make room in their theories, pedagogies, and practices for new digital writing spaces. Nine years have passed since the publication of Heidi A. McKee and Dânielle DeVoss’s groundbreaking examination of digital writing methods and ethics. Eight years have passed since McKee and James Porter grappled with the particular ethical challenges of digital writing research. What has changed since these foundational pieces were published?

In some ways, nothing has changed: pedagogical orthodoxies that privilege alphabetic text persist. On the other hand, much has changed in that digital research sites, questions, and methods have proliferated and digital rhetoric has become more firmly entrenched in the discipline, something reinforced by the job market postings that include digital rhetoric/writing.

We hope that these brief interventions offer ways of understanding how writing research methods and ethics continue to evolve in the face of ever-changing digital writing spaces and practices.

Megan McIntrye’s “#Tweet-Search: Perils and Possibilities” recounted her “experiences as a novice researcher thrown into digital research without much methodological training or technical skill” to explore the issue of informed consent when institutional review boards do not require consent.

Rik Hunter’s “I Wanna Be ‘RAD,’ But What’s the ROI?” explored the value of the concept of “return on investment” and the ethics of choosing more or less demanding methods and projects (i.e., un-“RAD” research) when navigating a heavy teaching load. In a publish or perish tenure system that seems to reward quantity of publications over quality—like getting a product out for sales—what are the ethics of publishing good-is-good-enough digital research?

Kerry Banazek, in “Objects to Think With vs. Objects of Study,” considered the interplay between materialist philosophy and empirical research’s relationship to materiality and described the traps
researchers sometimes fall into when discussing ethics and research design. She argued that there is no such thing as a research project that is not also a philosophical project.

Kelsey Cameron’s “Uncovering Web Histories” discussed strategies for accessing, collecting, and analyzing past iterations of contemporary websites and argued for the importance of attention not only to history but also the affordance and constraints of archival tools. That is, web history is not a single, static thing, but an evolving spectrum of approaches and technologies geared toward different ends.

Megan McIntrye: #Tweet-Search: Perils and Possibilities

In this presentation, I recount my own messy introduction to digital methods. My experiences as a novice researcher thrown into digital research without much methodological training or technical skill led me to ask three related questions: How do we best collect tweets? How/why code tweets? And is a profile a person? I argue that this last, ethical question is the most pressing of the three, especially for those of us interested in investigating digital protest movements or online harassment.

In April 2013, I had just defended my dissertation prospectus. My committee was happy with the shape of my theoretical argument, but they expressed concerns about my (lack of a) research site. Around the same time, an acquaintance from college began posting on Facebook about running the Boston Marathon. I had never known anyone who would run the marathon, and so, on April 15th, I watched the live feed of the race online. I was still watching when the bombs exploded. I quickly opened both Facebook and Twitter, trying to figure out what had happened and whether she was okay. As I awaited news of her whereabouts, however, I began noticing the multiplication of hashtags related to the race and the explosions. I watched in real-time as the official race hashtag became a place for sharing news, condolences, offers of help, and pleas for information about loved ones. In that moment, it clicked for me, and I realized this could be my site. If I wanted to examine ways that nonhumans participate in agency alongside humans, what better place to do so than in the middle of these storms of tweets?

The problem was, however, that I was not technically or methodologically prepared for that kind of project. I cobbled together methods and learned technical details and approaches on the fly, and reflecting on that experience now, I think I can offer linked questions (and some answers) to open up discussions about the kinds of methods and ethics that social media research demands.

First, though, I want to note that my interest in tweets is almost always qualitative. I find quantitative analyses interesting, but qualitative work allows me to contribute deep data to our emerging obsession with big data. Writing about her own experiences of research on Twitter, Alice Marwick (2013) made a similar point: “Qualitative research allows scholars to investigate the practices of a particular user group, as it can go beyond tracking follower counts or hashtag use to include many more sources of input about a specific community or user segment” (p. 109). Qualitative analyses of social media practices enable us to contextualize practices in ways too often elided by the meta-analysis at the heart of big data.

This interest in qualitative (or mixed methods) examinations of Twitter led me to my three questions.

First, a technical question: How best do we collect tweets? As I noted before, I was not technically prepared for doing the research I found myself most wanting to do. In the immediate aftermath of the explosions and the growing Twitter conversations, I had no clue how to even capture what I was seeing. So, I ended up with three 200+ page PDF screen grabs of various hashtag conversations—not the best way to capture, store, or set myself up for easy analysis later on. Depending on a researcher’s level of comfort (and preparation), tweets can be collected in a few different ways. In terms of low-tech approaches, one could follow my bad example and create PDFs or use a free, user-friendly site like HootSuite or Tweetdeck. For those with more tech comfort or for those willing to learn, it might make sense to create a Twitter developer account for free and then use a scraping program like DMI-TCAT or Gephy. If I had all the money in the world, I would have used a commercial service like Gnip, but since it costs upwards of $700 per month for access, that was not really an option for me.

Let me pause for a minute here to plug Bill Wolff’s step-by-step guide to programs for Twitter data collection, which is available at archives.williamwolff.org. I would also call attention to danah boyd’s
bibliography of Twitter research: www.danah.org/researchBibs/twitter, which includes various qualitative and quantitative studies that might be of particular interest at the outset of a project.

Next, I would like to pose a methodological question: Why code tweets? And how might we do it? For the why, I’ll turn again to Alice Marwick. In the same 2013 article on Twitter research methodologies, she said: “In providing thick description of specific tweets and interactions, we were able to illuminate specific patterns of use that would have been difficult, if not impossible, to ascertain with a more automated method” (p. 118). Coding, close reading, and rhetorical analysis invites this kind of deep examination and allows us to better describe the lifeworlds under examinations when we think about digital communities.

In terms of the how, researchers might treat tweets like other kinds of writing or verbal data: there is verbal data analysis (with its t-units and emergent codes), close reading, and rhetorical analysis. I have taken both these approaches with Twitter data and find the narrative potential in close reading and rhetorical analysis an important counterweight to much of the quantitative social media research.

Finally, I ask an ethical question: Is a profile a person? Two related concerns underlay this question: for one, is informed consent possible and what does informed consent look like in these spaces? First, a note: most institutional review boards do not require consent for public messages, like those on Twitter. Beyond that institutional responsibility, though, what are our obligations to those whose words we examine? As a number of researchers note, obtaining informed consent can be difficult or counterproductive, especially in situations where the researcher’s digital presence could interrupt the discourse community or when users operate behind pseudonyms (Beninger, Fry, Jago, Lepps, Nass, & Silvester, 2014; Woodfield, Morrell, Metzler, & Blank, 2013). On the other hand, using profiles without contextualizing their content, something that informed consent makes more likely, may lead to inaccurate or incomplete analyses (Zimmer & Proferes, 2014; Beninger, et al., 2014). I find these questions especially pressing for those of us interested in digital protests movements or harassment online. Both of these practices require balancing sometimes conflicting sets of concerns or allegiances.

Perhaps, then, I might reimagine this last, most pressing question: How do we balance privacy concerns with the value of critique?

**Rik Hunter: I Wanna Be "RAD," But What's the ROI?**

In this presentation, I explore the value of the concept of “return on investment” (ROI) in considering Haswell’s call for more support from our leading journals for replicable, aggregable, and data supported (RAD) research in composition. How has ROI become a deciding factor in the projects I pursue and the methods I use pre-tenure? Working in a teaching-intensive institution, where publishing expectations have (reportedly) risen over time, I explore an admittedly and deliberately provocative and mercenary-like attitude in suggesting ROI and methods and methodologies should be discussed with graduate students and newly-hired junior faculty seeking to earn tenure.

I love Shark Tank. I love seeing passionate people coming up with creative solutions to problems, making arguments for their creations, and talking about audience. The pitches are practiced and rhetorical, as are the products and services being offered . . . hopefully.

Beyond the rhetorical interestingness, another aspect of Shark Tank that I find fascinating are the discussions of return on investment (ROI). That is, it does not matter how well your product solves a real problem for people. If the time and money invested into it does not return a great enough profit, then you have to, in the words of Mr. Wonderful, “take the idea out back and shoot it.”

Even more illustrative of the ROI mentality, Barbara Corcoran (2016), Shark and real estate mogul, recently told one inventor,

> This is what I’m really disturbed by, I think you’re the type of person who is careful plodder, so I don’t believe that you’re the kind of entrepreneur that’s really going to get the product out there … for sales.

Her statement—regarding a certain type of entrepreneurial ethos—aligns nicely with a talk given by Andrew Kirplani (2011), a software engineer and CTO at WorkHound. When he spoke to a group of
software engineering students at the University of Iowa, he argued that engineering students need to be more like business students. You’ve got to stop striving for perfection before releasing software because your software is only worth what someone will pay for it.

1. Business is about money.
2. Unreleased products don’t make money.
3. Good is good enough.

That is, software engineers are in business selling time and effort to get a paycheck. Stop trying to make it perfect. See Windows VISTA for example!

![Screenshot from Andrew Kirplani’s talk on the differences between Business and Engineering majors.](image)

It may be a leap, but I think we can make strong connections between what we do and this idea of ROI. At the very least, it is a conversation we should be having, especially with junior faculty, with regards to our research methods, the size of projects, and how great a ROI we can generate on some projects over others.

And I ask this question as someone who has invested a lot of time into learning about a variety of research methods and methodologies. I have made efforts to be RAD. For instance, I took several methods courses during my doctoral studies. I did the two-week Dartmouth Summer Seminar and was designing a project that I ultimately think does not offer a great enough ROI. And, I have more recently had my mind blown by Bill Hart-Davidson and Ryan Omizo in their full-day workshop, Computational Rhetorics.

So, it is not like I am anti-RAD. There is a part of me that loves counting T-units.

But beyond my vacillating between my right-brained affinity for more-narrative methods and what we might call hard research, I am also quite aware of the context in which I work every day—a teaching-intensive mid-size regional with increasing pressure to publish.

In my 4/4 world, ROI has become a deciding factor in the projects I pursue and the methods I use because some of my literature colleagues admit to not getting comp-rhet scholarship. A bit hyperbolically, I am essentially adding publishing bullet points to my CV.

Perhaps that will change after tenure. For instance, in the recent Chronicle essay, “Failing to Dare Greatly,” Rachel Toor (2016) wonders “if the current system and climate in higher education prohibits anyone but the comfortably tenured from being allowed to fail while daring greatly in their scholarly work.”

“Conservatism,” she writes, “is certainly the easiest path.”

This conservatism, and the pressure to publish, seems to also create a problem in those disciplines that are necessarily RAD.

For example, just over a week ago, Daniel Sarewitz (2016), co-director, Consortium for Science, Policy & Outcomes Professor of Science and Society, School for the Future of Innovation in Society, wrote about how the pressure to publish makes it more difficult to find good research.
It is hard to imagine a place where research could be more RAD, and yet, poorly done studies, rushed to meet publishing expectations for tenure, is leading to worse scientific knowledge, not better.

In my context—and perhaps we can talk about all of your contexts—it seems that you could follow this probably cynical formula that I have created to increase your scholarly ROI (and maybe reduce your level of stress):

- Publish un-RAD research PLUS
- Publish good-is-good-enough research PLUS
- Publish it often PLUS
- Do not dare too greatly and be conservative EQUALS

Tenure, with a greater return on investment.

**Kerry Banazek: Objects to Think With vs. Objects of Study**

This presentation invites consideration of the interplay between materialist philosophy and empirical research’s relationship to materiality. I give a hyperbolic introduction to three traps researchers sometimes fall into when discussing ethics and research design. In particular, I suggest there is danger in accepting that what is at stake when conversations about philosophy and research methodology come together always has to do with epistemology, never ontology.

I spend a lot of time thinking about the question: What happens to the idea of ethics in the context of new materialist philosophies? And I like to claim: materialist conversations that have been going on for decades within composition and rhetoric—including the conversations about digital, data-driven research methods my co-presenters engage—can be used to ground that huge, unwieldy question.

Consequently, instead of detailing a research project, I am going to use my portion of this roundtable to belabor something most people in the computers and writing community probably take for granted: there is no such thing as a research project that is not also a philosophical project. We get an inkling of how true this is every time one of my co-presenters uses the word *ethics*. We can also associate this fact with the frequent winking that links a broad material turn that has swept across the humanities and social sciences with an ontological turn in rhetoric (a linkage that suggests the work of Scot Barnett, Casey Boyle, Diane Davis, Thomas Rickert, and many others).

And yet, projects that engage explicitly and thoughtfully with both philosophical and empirical traditions of writing and media studies remain rare. We could chalk that up to the fact no one can be an expert in everything and move on. But I am suggesting we might not want to do that. Especially if we care about realisms and materialisms, which empirical researchers tend to. To that end, I want to offer three snapshots of ways of thinking to help all of us think through what is going on underneath my abstract claims and what might be at stake.

**First**

A snapshot of a nasty trap that can be triggered just by letting the words research, methods, and ethics get kind of near one another.

Someone makes the seemingly innocent claim: good research is systematic. Maybe that just means careful, we think. But then we wonder if systematic could be a synonym for empirical. Then, via a slippery slope we start imagining empirical methods are the most ethical methods available. If we do not like saying ethical, we can say instead, empirical methods are the most rigorous methods, but that does not change what is happening. We still get to be smug. This is realist, rationalist, Enlightenment thinking at its best.

**Second**

A snapshot of a reaction that is its own kind of trap.

Perhaps because of the aforementioned smugness, it is oft-implied that there has always been a simmering conflict between empirical and non-empirical researchers in composition. This is a general claim
but also a claim Ellen Barton made explicitly in 2000 in an article questioning a kind of mass ethical turn in composition field research. The work she associated with that phrase drew on feminist ideas; it steered way wide of the above trap, emphasizing development of personal relationships between researchers and subjects. Along the way, it insisted that studies become ethical by having reflexivity and reciprocity built in.

That is mostly awesome, Barton notes, but adopting these values wholesale in a moralistic way pushes us to conduct small-scale research. By discouraging the kind of large-scale work where personal relationships might be impossible to build, it keeps us from asking certain questions and seeing patterns; that is problem enough, but moralizing of methods also forecloses conversations about how complicated ethical problems are. And it makes thinking about scale change nearly impossible. Basically—and these are my words not Barton’s—being dumb about research methods makes us dumb about how ethics work, too. There is no such thing as a methodology that is always ethical.

Third

Some quick and dirty claims.

Both of these traps are made more enticing when we try to pretend that we can have preferred methods without our philosophical opinions about the nature of the world leaking all over. More specifically, we get ourselves in trouble by accepting that what is at stake when conversations about philosophy and research methodology come together always has to do with epistemology, never ontology.

It is easy to see where that assumption comes from; ontology is concerned with being and existence. Epistemology is concerned with knowledge. Given the chance, many people would rename ontology as stuff that might be true but we can’t know, so worrying about it is a waste of time. They would rename epistemology as stuff we know that can influence people’s actions. Why would we not stick to the latter? Epistemology is not just the question: what is the nature of Knowledge-capital-K? It is also: how do we know [x], when [x] can be anything? And also, where is this knowledge instantiated? How is it distributed? Who knows [x]? Who does not? It encompasses all kinds of questions that ought to be asked during research design. The impulse to focus on related questions speaks to Jeffery Bineham’s observation that concerns characteristic of the “epistemology debates” provoked by Robert Scott, which helped shaped the field of rhetoric in the 1970s and 1980s, remain central to the field but have been disassociated from the term epistemology, “translated into concerns for how rhetoric influences ‘inquiry’” (p. 43). A quarter century has passed since Bineham made that observation, but if anything, it holds more strongly now because so many have forgotten that an act of translation even took place. Never having to say epistemology out loud means we never have to say we are avoiding talk about ontology, which makes it easier to get away with.

But when we bracket off ontology, we limit ourselves. Choosing to bracket ontology threatens to cut us off from the process of invention. If we have stuff we know on one side of a balance, and stuff that is unknowable on the other, then where does stuff that is knowable but that we don’t know yet go? Where do we put speculations that are grounded and carefully composed but which have not achieved the status of fact? Where do we put things that our bodies know that we can not explain? How do we develop strategies for living with new-to-us phenomena?

Moreover, if we bracket ontology, we lose the ability to be realists in any meaningful way. This matters to me because the big lesson I have taken from the weird, speculative, materialist philosophies that I spend a lot of time with is this simple old materialist idea—that there is a world out there means: we can get things wrong. And we do. All the time. We cannot help it. Not coincidentally, without crediting ontology explicitly, structuring exigencies can become an unnecessarily convoluted struggle. If you do not believe we can get things wrong, why would you bother doing research, especially empirical research that is driven by experience and data? Sure, the word ontology is pretentious and specialized, but the two questions that fill out the ontological realm are just: what is there? and what is it like? And humans, all of us, spend a lot of time fiddling with these questions. Which loosely implies my last and most contentious complaint—the act of bracketing ontology is part of what separates theory-minded-humanists from the rest of the world. For worse, not better.
Kelsey Cameron: Uncovering Web Histories

In this presentation, I discuss strategies for accessing, collecting, and analyzing past iterations of contemporary websites. Using examples from a project on the evolution of sites for queer women, I argue for the importance of attention to history—and particularly histories of writing and revision—in approaching a medium often aligned with the present.

We often imagine that digital technologies fix problems of preservation. For example, a Dropbox ad promise that digital file storage will safeguard your things against the fragility of analog existence: “Any changes can be undone,” it proclaims, and even deleted files can be “undeleted” (Ryan Singel, 2011, n.p.). This ad fed into a larger narrative equating the digital with perfect memory. As Wendy Chun (2008) glossed it: “Digital media, through the memory at its core, was supposed to solve...archival problems such as degrading celluloid or scratched vinyl” (p. 153).

The reality, of course, is less utopian: even as digital archives proliferate, the digital creates its own unique archival problems. The one I am going to talk about today has to do with web history—more specifically, with how hard it is to access websites from the past. This is an endemic issue, built into how the web works: As Steven Schneider and Kirsten Foot (2004) write, “Unlike any other permanent media, a website may destroy its predecessor regularly and procedurally each time it is updated” (p. 115). We are aware of some of these acts of destruction: consider the outcry the last time Google retooled its logo, or when Twitter switched its favorite icon from stars to hearts—there is a felt, much remarked upon difference when popular sites tweak the images and interfaces we have grown accustomed to.

But eventually we forget the changes; on less trafficked sites, we may not notice them at all. Thus, elaborate histories of writing and revision disappear beneath a website’s surface. In the rest of my time, I am going to detail some reasons we might want to get at those histories and some methods for excavating them.

So, first: why should we care about web history? There are a lot of potential answers here: interest in web design and its historical changes, insight into what kinds of web artifacts best endure, and access to particular community histories embedded in websites. My project, which is about the emergence of contemporary websites for queer women, comes out of this last impulse. Two sites—Autostraddle (https://www.autostraddle.com/) and AfterEllen (http://www.afterellen.com/)—were my main focus, and my goal was to track how they evolved: when they scaled up from small, individual projects to community spaces, and what precipitated those changes in scale.

Given the lack of institutional memory about queer things, artifacts like websites and online forums can provide traces of the past we have no other way of approaching. Thus, I see the update logs of AfterEllen and Autostraddle as windows into queer cultural history. So, now, on to the methods.

My first stop was the Internet Archive’s Wayback Machine, a huge, public collection of historical web data assembled using Alexa Internet’s web crawler (https://archive.org/web/). The Wayback Machine imagines the website as a unit of online history: you input a URL, and then you browse through the various iterations of it in the archive. This works better for some sites and for some projects than others. In my case, it was a useful chronicle of AfterEllen’s development: the archive contained many snapshots of the site captured from 2002 to 2016, and by looking at them I could see its gradual shift from grassroots organization to corporate media structure after it was bought by Logo in 2006.

There are aspects of web history that escape the Wayback Machine’s archive, however. In researching, I encountered three main limitations to its utility.

Limitation 1: The Trouble with Robots

Sites that do not accept robots—that cannot be crawled—do not get archived. For me, this meant no access to OurChart, a lesbian social network run by Showtime from 2006–2008. Though this site was a focal point for queer female community during its lifespan, no trace of it appears using the Wayback Machine. Things that are not accessible in this way are not necessarily gone, however: I found images and a description of OurChart through Albertson Design, the company Showtime paid to create it. This
discovery speaks to how commercial and industry discourse can help fill in archival gaps: if a web artifact was commissioned, there are likely publicity materials about it posted somewhere.

Limitation 2: The Trouble with Images

The Wayback Machine archives images irregularly, so the accuracy and completeness of a site’s visuals are questionable. Because of the way it composes pages, it also sometimes makes recreations that never actually existed, combining images and content from different crawls into never-before-seen versions of a site.

Limitation 3: The Trouble with Scale

While you can sometimes use links to move between pages, the Wayback Machine does not chart the topography of past cyberspaces. So, if you are interested in content clusters—say, what it looked like to Google lesbian social network in 2006—it cannot get you there. It offers access to individual sites but not their interrelations; you get little sense of what happens at scale and the modes of navigation available to past web users.

These limitations are specific to the Wayback Machine. Other projects are creating different sorts of internet archives that can accommodate different research questions: there are the Library of Congress’s thematic MINERVA collection, national efforts in Australia, Denmark, and elsewhere, and—perhaps most relevant to this panel—the WebART project at the University of Amsterdam geared toward data-based humanities research. These efforts remind us that, though the Wayback Machine may be the most visible and most popular web historical tool, it is not the only one through which we can explore the internet’s past. We should not let its structure dictate our research trajectories, for web history is not a single, static thing, but an evolving spectrum of approaches geared toward different ends.

References


Digital Research Methods


Woodfield, Kandy, Morrell, Gareth, Metzler, Katie, & Blank, Grant. (2013). *Blurring the boundaries? New social media, new social research: Developing a network to explore the issues faced by researchers negotiating the new research landscape of online social media platforms*. London, UK: NatCen Social Research.