Week 2.1 Reflections

- Rosenberg, Daniel. 2013. "Data before the Fact." *Raw Data Is an Oxymoron*, 15–40.
- Ramsay, Stephen. 2014. "The Hermeneutics of Screwing Around; or What You Do with a Million Books." In *Pastplay: Teaching and Learning History with Technology*, edited by Kevin B. Kee, 111–20. Ann Arbor: University of Michigan Press.

Anya Kalogerakos

1:06 PM Both readings for Tuesday seemed to touch on the issue of having too much information. In Ramsay's article, he draws a distinction between the practice of searching and browsing, one of which (browsing) has been somewhat lost in the advent of the web. While I enjoyed reading his take on how to handle the increasing amount of knowledge being added to the human canon, I do think that he missed a large part of Internet searching which does include serendipitous browsing (aka doom scrolling, getting sucked into a rabbit hole, following way too many links via Wikipedia). However, I won't hold this against him as the Internet probably looked different at the time he was writing this. Rosenberg, on the other hand, draws a distinction between data, fact, and evidence, tracking the evolving use of data throughout the past couple of centuries. From his writing, it is interesting to think about what a world without the concept of "data" (as we know it today) would have looked like. Rosenberg's discussion of Priestley's work was also interesting to think about, as the organization of the data was the innovation, as opposed to the data presented. In our current time, the organization of data is becoming almost (if not more) important than the data we have being collected. Data is everything now, but at the same time, meaningless without some sort of structure.

Pia Bhatia

2:33 PM What struck me most about Ramsay's article was how seriously he treated stumbling upon interesting books as a central aspect of study within the humanities, and by no means suggested that this aspect of the process be eliminated or made more 'efficient' using softwares. I was additionally struck by his characterizations of categorizing books – there was a quote where he suggested grouping them by color, but the idea was developed beyond that, leading me to wonder if applications would someday alter our understanding of how a book is described (what the 'Classics' are, for example), that various literary canons may be unrecognizable from what we once considered them to be. I wonder what would emerge as popular, literary, etc. and how the use of geographical regions in taxonomizing literature would be affected.

Alison Fortenberry

4:19 PM The Rosenberg article was really interesting because it forced me to think about how I define data and how that is shaped by my environment. Even though I've used the term "data" to represent pieces of evidence in the humanities, I still tend to conceive of it as something kind of inherently technological and STEM-related. The reframing of data as a rhetorical concept is a really helpful way of contextualizing why I conceive of data in that way, and understanding why historically people have understood it differently. The Ramsay reading was also really interesting, especially in the idea of searching versus browsing. I think in a lot of ways, it's harder to stumble across something new now because of the optimization of online search engines. I've also found that intentionally trying to browse online is pretty difficult with targeted advertisements. If I want to look for a new book in a real bookstore, I can go to multiple different sections of

the store and find interesting books from different genres, authors, and interests, making it easy to discover something new. In a digital bookstore, however, it's a lot harder to do that. I can search by genre, but it's generally a curated list of books I may enjoy based on what I've already purchased and enjoyed. A lot of attempts at digital browsing feel less like self-piloted browsing and more like advertiser-driven searching on one's behalf.

Clay Glovier

4:56 PM I found Rosenberg's article on data very interesting. I never really thought about the differences between the words data, fact, and evidence, but I agree that they each serve a uniquely distinct purpose. While facts cease to be considered facts once they are proven false, Rosenberg shows how data considers to be data, just "false data." Thus, the term is more rhetorical than factual. Ramsay's article was also fascinating. One interesting part was his discussion of Gregory Crane's imagination of a library, where he said that the purpose of millions of books was to enable people to "extract from the stored record of humanity useful information in an actionable format for any given human being of any culture at any time and in any place." I was doing just this as I searched the terms used in this article, many of which referenced people and events I had not encountered before. I was finding relevant pieces of information which helped me in real-time to understand the article about that very activity. It was interesting how Rosenberg said Google Search was not well suited for browsing. I agree, you can find what your looking for and that's it. However, I am not sure how one could design it so that you could get from gardening as he says to creation myths easily. If you search for gardening tips, that is what you are looking for. If a bunch of other material comes up that could be distracting, but maybe it could be placed lower down in the search so that you could find what you are looking for and then explore other related interesting topics. (edited)

Melissa Woo

5:24 PM What strikes me about both of these readings is the extent to which they represent the effect that humanities can have on technology. With a more technical background, my initial thoughts with regard to applications of digital humanities are based on applying technical methods to humanistic datasets or open questions. For example, one could study linguistic patterns by visualizing data over time, or use word embedding vectors sourced from textual mining to investigate literary trends. Rosenberg analyzes the early history of the concept of "data" to consider how the space was formed. Ramsay considers of the "hermeneutics of screwing around" to conclude that one's ethical obligation to the vast network of available information is not to "read them all or pretend that you've read them all, but to understand each path through the vast archive as an important moment in the world's duration – as an invitation to community, relationship, and play" (119). Rosenberg's work impacts how we understand the nature of data – whether we take it to be factual or not, whether we use it as evidence to support an existing argument or as an exploration to find a new argument, whether we perceive it as conceptual or practical. Ramsay's work could have very practical implications for the architecture of web browsing and information search from vast networks, changing the way in which we find, disseminate, and interpret information. Both of these theoretical and humanistic considerations have tangible implications for technological pursuits.

James Sowerby

11:32 PM I really liked the readings this week, and felt like they complemented each other well. Ramsay's key distinction between searching and browsing was helpful for me in my own work to realize the limitations of my research habits. What I tend to do—and what has been emphasized to me throughout both high school and college—is this kind of specific searching. I don't tend to think about how, as Ramsay writes, I

am "left with a landscape in which the wheel ruts of [my] roaming intellect are increasingly deepened by habit, training, and preconception" (Ramsay 118). Already this article is productive in the sense that it encourages me to remember the serendipitous—the "screwing around" or the "amusing geegaw" potential of the Ngram viewer." I do wonder, however, how this functions in an environment like Princeton. How does one lean into this other side of research when there are real consequences when no such interesting connections are made, and the deadline is fast approaching? Rosenberg also had a really interesting discussion of truth and facts. In much of the work that I have done so far in the Humanities at Princeton, I don't recognize my sources as data per se. I often have assignments that are close textual analyses and so I work mainly off of the authors' written word. But sometimes the passages at hand can have really interesting relationships to truth. Insofar as the written word is data, the do not necessarily always contain truth—in the case of an unreliable narrator, for example, the data are based off of his words. It is not false that he says one thing, but rather becomes false in the context of his continual contradictions throughout the rest of a novel. This is, perhaps, another step in applying these lessons to my own life.

Pippa LaMacchia

1:19 PM These two readings were fascinating to read in tandem because they explain two different types of data-gathering. I was particularly interested in Stephen Ramsay's article because as an avid reader I often sit with the discomfort of knowing I will never read every book. I have never tied this question to the digital world but hearing Ramsay make the connection (and distinction) between browsing in a tangible library versus browsing an online library was fascinating. It makes me question how much autonomy I have online (not much, I know that already) but it also makes me want to discuss the notion of 'randomness' and spontaneity. So much of what people (humanists) uncover or research begins with coincidence or, as Ramsay calls it, "serendipitous engagement," that I have to ask if the technological world is a detriment to the human experience of curiosity or if it helps it along. By then going to Rosenberg's article we might be able to connect the idea of "data" to this too-structured online sphere because data has no right or wrong (i.e. "data has no truth"). I wonder if the data we gather from randomized in-person browsing is equal in value to the data we gather online.

Andrew

10:45 PM Ramsey's article was the most interesting of the two since it touched upon humanity's drive to consume as much knowledge as possible in the limited timespan of our lives, and moreover, questions about canonicity and the quest to access the best form of "condensed knowledge." My first question is: why do we have this urge to consume the perfect amount of educated information? I believe that there comes a point where the goal outweighs the potential of people's intelligence and brain capacity to the point of over-saturation. Perhaps it is a good thing that professors have limited knowledge or that we specialize in understanding a few aspects very thoroughly (film, literature, data, philosophy, etc). Next, the question of defining the canon is very interesting, and I agree with Moretti that there is no "practical solution." I think having a defined canon of education limits the potential and route for individuals to expand their understanding in unique ways that touch them. It is this diversity of thought that will better humanity. Einstein said, "The only thing that you absolutely have to know is the location of the library." (from the Tiger Tea Room in Firestone); thus, as long as there is a manner of interconnecting all facets of information through technology and teach the world how to access and use this technology easily, knowledge will be able to spread in a more organic and personal way. The modern age of over-saturated content and information is currently too overwhelming and detrimental for the passing down of true/real information, and media overall has a negative effect on individual thought and expression. Lastly, I find it interesting that

Rosenberg makes the distinction between data and fact, saying that in the modern age, “data has no truths.” The hyperinflux of information has made it very difficult to question their veracity. In the end, I agree with Ramsey’s conclusion and Gregory Crane’s vision that we need an easy way to “understand each path [of information] through the vast archive” and, on top of that, somehow reinforce truth.

Colin Brown

11:04 PM One of the first patterns that I noticed was the central role that Google plays in both of these readings, not to mention the Ngrams activity. It seems to really emphasize that the ability to aggregate, search, and connect massive collections of data has profound implications for how the humanities study their field of work. I also found interesting the contrasting perspectives on Google: Rosenberg acknowledges its power but ultimately says it is incorrect for the information he looks at, while Ramsay applauds how using Google is like a supercharged version of browsing in a library and finding new, useful information along the way. Rosenberg talks about it, but I would be curious to see Ramsay’s views on other online information databases. My biggest moment of insight came from Rosenberg’s timeline of the evolution of the term “data”. I felt like the lexiconical story perfectly aligns with how data is viewed today - it is observation of some kind, not necessarily assumed truth. This seems important today given how widespread misinformation and deep fakes are these days. It was almost heartwarming to hear that data was one regarded as innate facts; I wonder if the pendulum will ever swing back that way.

Helen Gao

12:43 AM The history of the word “data” described in “Data before the Fact” was very interesting, and I especially liked the idea that “False data is data nonetheless”, which challenged the idea of data as truth while still emphasizing its importance to our modern life. I think with the recent hype around the different uses for data, such as using it to extrapolate additional information, this is an especially important clarification to make. I also liked how the article introduced the Ngram Viewer tool that we used this week and provided examples of word usages that reflected historical changes. “The Hermeneutics of Screwing Around” was also quite interesting, although I personally found some of the author’s arguments about browsing on the Internet to be a bit outdated; I think that the modern Internet does an equally good job of facilitating browsing as the traditional library, and with many times the amount of available information. Still, the article made an interesting distinction between searching, which has purpose, and “screwing around”, which browsing, which is more aimless. It made me think about how I personally use the Internet to do both – I spend significantly more time browsing on various social media sites, but I do use search engines to search on a regular basis as well.

Layla Williams

12:56 AM This week’s readings made me rethink the importance of choosing my words carefully (not only in the future when using search engines) when discussing any data in the course. When I thought of data in the past, I thought something about the word made it inherently correct. In other words, by its nature, data was always to be trusted and correct — because it was data, of course. However, it is more helpful to understand data as something that we must then interpret, interrogate, and learn from any discrepancies and mistakes within it. Thinking about data as the words within a novel and the analysis of the data as how I might close read a novel has also helped me understand some of the goals of the digital humanities a bit more as well. However, I think we might also run into some trouble in trying to transfer slightly more creative work in the humanities directly into data. For example, how is the context of language within, novels for example, taken into account when analyzed through digital means? How are methods developing to

understand the emotional nuances of how language might be used within a novel or play? Might these methods be more beneficial for browsing rather than searching for a specific work?

Talia Goldman

1:02 AM I particularly enjoyed Ramsay's "The Hermeneutics of Screwing Around," which strengthened my understanding of how digital resources have affected research practices in the humanities. Clearly articulating the potential benefits between online browsing alongside the pointed aims of an Internet search, Ramsay highlights the deficiencies of digital tools, which do not always support browsing-style learning, in aiding with productive research. I agree that the loss of serendipity in research limits creativity and may lead researchers to go in circles when researching, and would like to see more "screwing around" enabled when conducting research or simply following curiosity online. That said, I am not sure how this is possible in research databases and DH-specific tools, aside from the existing "you may also like..." options or hyperlinks that lead readers to new, but often very similar, material. The Rosenberg reading also provides interesting insight into how DH tools relate to research by chronicling his research into the history of the word "data," noting that we need to attend with greater care to the "epistemological implications of search" (Rosenberg 35). Together, these readings emphasize some of the deficits of DH tools in how they guide and determine paths of humanities research. From this, I take away the need to be critical of DH tools, which (like the Google Ngram viewer) may produce ambiguities and statistics not to be taken as fact, but, as Rosenberg notes, as data that may morph into fact and evidence, but is not fact nor evidence in and of itself. After considering these two readings, I find myself engaging more critically in distinguishing data from fact, thinking more deeply about how published research (or even posts on social media, which are especially tricky in its use of data) place particular, but perhaps not exhaustive, meaning onto data sets.

Yaashree Himatsingka

2:02 AM In an age of 'information overload' (though the readings reveal that previous ages were consumed with their own versions of this phenomenon), I've definitely wondered about how to navigate knowledge in an ordered and coherent (also achievable, sigh) way – especially as a History major. I appreciated Ramsay's acknowledgement of the power dynamics that underlie the processes of classing and curating knowledge, especially his emphasis on randomness and his problematization of discernment (i.e., "...no one can really say what is "representative" because no one has any basis for making such a claim.") In Rosenberg's reading, I was struck by the points made regarding the fallibility of data ("...the existence of a datum has been independent of any consideration of corresponding ontological truth. When a fact is proven false, it ceases to be a fact. False data is data nonetheless.") This underscores the complexity of knowledge management in a time when data is abundant but not always reflective of truth. What degree of skepticism do we employ, for instance, while using tools like the Ngram Viewer? (I wonder about which books Google has digitized, and how many editions of each? Is there a bias towards period or place?) (edited)

Emanuelle Sippy

2:18 AM I am interested in the way these pieces tease out the relationship between data, ontology, and epistemology. In particular, I think Rosenberg makes a compelling case for the possibilities and limitations of these tools. He advocates for using them with a critical disposition that we need to integrate more fully into educational spaces where these tools are being used more and more frequently. Rosenberg's point about a healthy skepticism is echoed by Ramsay when he describes the interplay of "cultural anxiety" and "cultural confidence." I am wondering about the ways in which we can instill this balance in educational spaces, which would better position both students and scholars to use these tools with greater intention and

accuracy. With the rise of technological tools for research and everyday use, we seem to have an even greater need for intellectual humility, and yet our societal impulse seems to be more opposed to that humility with the production of each tool. How do we address this? Are there ways, perhaps, even within the tools themselves, to embed information that inspires personal and collective humility? Is there information about the data these tools are using (such as what Rosenberg details) that should be more obviously available (i.e. noted in the form of disclaimers), especially for users who are not scholars and are not looking to (or may not even have the skills & tools to) corroborate the results? (edited)

Raphaela Gold

2:48 AM Before reading the Rosenberg article, I had never before considered the significance of data being not just pockets of individual fact, but rather “constellations of information.” Rosenberg continued to explain the significance of data in groupings throughout his piece, explaining the etymology of the word as plural from Latin, despite our modern comfort with using it to refer to the singular. Overall, I was very curious about Rosenberg’s use of etymology to make his argument, and it made me wonder how the field of linguistics in general intersects with that of digital humanities. For example, when I read that data comes from Latin “datum”, my first instinct was to take this as an indication that people consider data a gift – the gift of information and the potential for corroboration. Instead, Rosenberg took it in an entirely different direction. Rather than a gift, something people are generally grateful for, Rosenberg claims that the etymology of data paints it as a given in an argument, something that can be taken for granted and upon which a greater theory can rest. Regardless of the interpretation of data’s origins, I wonder how the ambiguity of language might contribute to or detract from our ability to trace the data of data. This sense of ambiguity seemed at odds with the concept Rosenberg raises later in the article about the reliability of data – how data is often used to identify categories of facts and principles that are beyond argument. But if we take data to be “beyond argument”, how can we remain in a position to constantly question and challenge it? And how to we define “beyond argument” in an age when there is disagreement even with regard to facts? Additionally, to quickly touch on the Ramsay article, I was struck by the debate he raised surrounding canonicity. Ramsay asked the question of what one does with a million books, but before that, we must ask the question of how to choose which books make it onto that list in the first place, and whoever makes that choice ultimately has a great deal of power in determining canonicity.

Ethan Haque

7:20 AM I firmly believe that most published research (and even more unpublished research) is all made up, lies, and doesn't reflect anything more than a fantasy world constructed by the authors. Some huge percentage of scientific works are necessarily incorrect. Conflicting viewpoints are not just common. They are sometimes necessary so that a researcher can say something interesting and get a job or for a professor to get tenure. We are incentivized to say bold things and things that may not be objective truths if only to have a unique opinion that is worth listening to. I think this is what, in part, contributes to the treatment of data as objective truths instead of biased distillations of some observation. In order for any valid argument to hold the data underlying it must be unimpeachable. This is almost never the case, though. Rosenberg puts this sentiment in a compact form at the end of the chapter stating, "data has no truth." As we get further away from the more scientific disciplines and move towards sociological studies this issue gets compounded. It gets harder to see what data is bad and what data is good.