Digital

[humanities]

[science]

[literacy]

Amelia McNamara

Digital humanities in the framework of data literacy

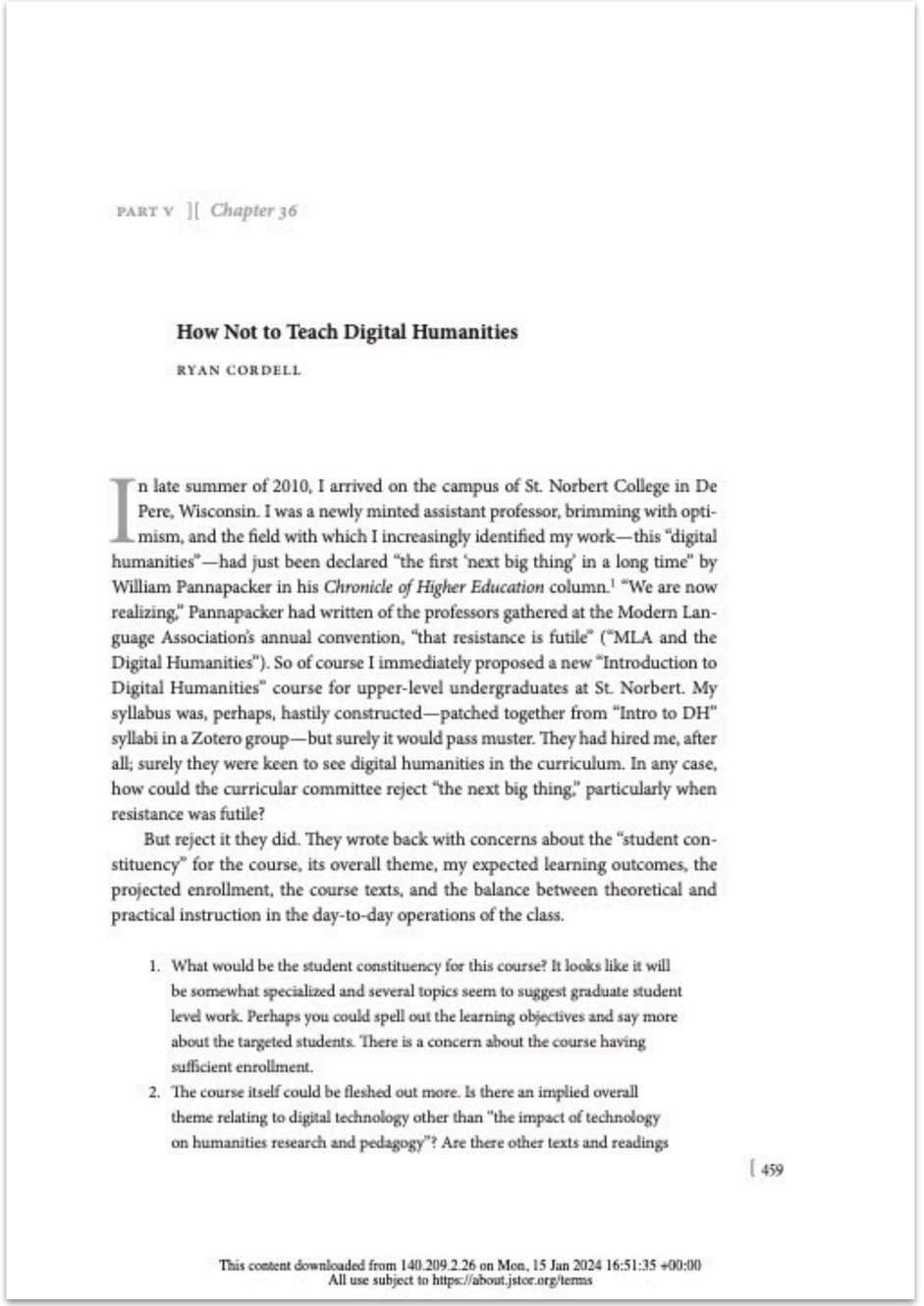
Science, Technology, Engineering and Math (STEM) often already include a lot of data/digital literacy

Humanities, Arts, and Social Sciences (HASS) are becoming more data-driven disciplines

“Digital humanities” is a useful search term, although it doesn’t capture everything. The digital humanities use digital tools toward the goal of advancing humanities knowledge. These digital tools might be distinct from data, but many aspects of digital humanities rely on data.

But…

- “What Is DH?” Always Excludes
- “Humanities” Is a Vague and Often Local Configuration
- Undergraduates Are Scared by Digitality
- But Don’t Panic
  - Start small
  - Integrate when possible
  - Scaffold *everything*
  - Think locally
- Whither “Digital Humanities”?



How Not to Teach Digital Humanities

RYAN CORDELL

In late summer of 2010, I arrived on the campus of St. Norbert College in De Pere, Wisconsin. I was a newly minted assistant professor, brimming with optimism, and the field with which I increasingly identified my work—this “digital humanities”—had just been declared “the first ‘next big thing’ in a long time” by William Pannapacker in his *Chronicle of Higher Education* column.[1] “We are now realizing,” Pannapacker had written of the professors gathered at the Modern Language Association’s annual convention, “that resistance is futile” (“MLA and the Digital Humanities”). So of course I immediately proposed a new “Introduction to Digital Humanities” course for upper-level undergraduates at St. Norbert. My syllabus was, perhaps, hastily constructed—patched together from “Intro to DH” syllabi in a Zotero group—but surely it would pass muster. They had hired me, after all; surely they were keen to see digital humanities in the curriculum. In any case, how could the curricular committee reject “the next big thing,” particularly when resistance was futile?

But reject it they did. They wrote back with concerns about the “student constituency” for the course, its overall theme, my expected learning outcomes, the projected enrollment, the course texts, and the balance between theoretical and practical instruction in the day-to-day operations of the class.

1. What would be the student constituency for this course? It looks like it will be somewhat specialized and several topics seem to suggest graduate student level work. Perhaps you could spell out the learning objectives and say more about the targeted students. There is a concern about the course having sufficient enrollment.
2. The course itself could be fleshed out more. Is there an implied overall theme relating to digital technology other than "the impact of technology on humanities research and pedagogy"? Are there other texts and readings

"How not to teach digital humanities.” Ryan Cordell (2016)

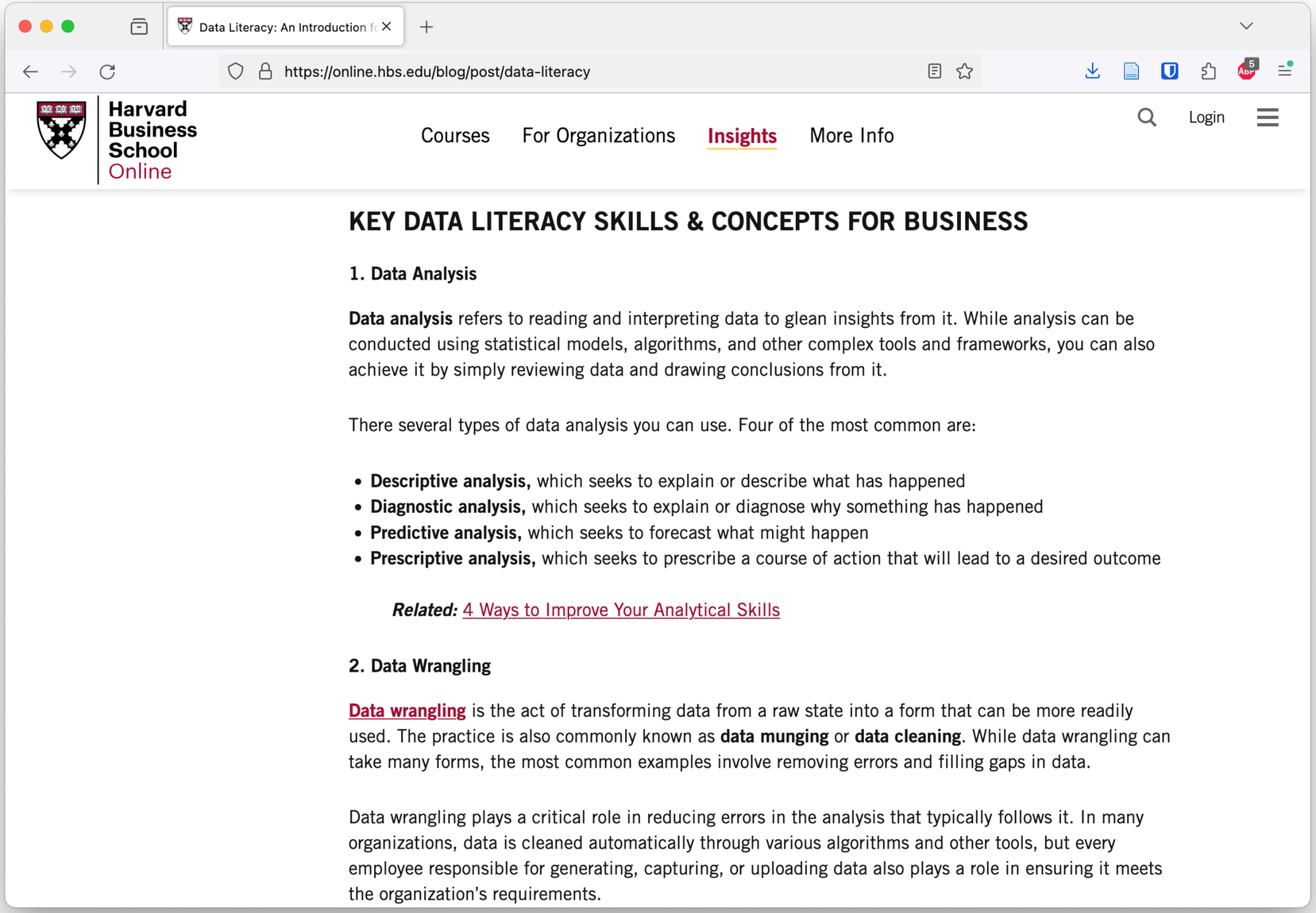
rondiorio
@rondiorio

Follow

#newsrw whether data journalism is journalism reminds when we argued about whether digital was real photography

3:52 AM - 27 May 2011

1. Data analysis
2. Data wrangling
3. The data ecosystem
4. Data governance
5. The data team



Data literacy, an introduction for business. Tim Stobierski https://online.hbs.edu/blog/post/data-literacy

- Data Concepts and Culture
  - Data Culture
  - Data Ethics
- Reading
  - Data Discovery
  - Evaluating and Ensuring Quality of Data
- Writing
  - Data Collection
  - Data Management and Organisation
  - Data Manipulation
  - Data Curation and Reuse
  - Metadata Creation and Use
  - Data Conversion (Format to Format)
  - Data Governance
- Comprehension
  - Data Analysis
  - Data Interpretation (Understanding Data)
  - Identifying Problems Using Data
  - Data Visualisation
  - Presenting Data (Verbally)
  - Data Driven Decision Making
  - Evaluating Decisions / Conclusions Based on Data

Data to the People https://www.datatothepeople.org/

DATABILITIES®

Databilities® is the world's first, evidence-based data literacy competency framework. Following an extensive review with Industry Leaders and Data Literacy Subject Matter Experts, the framework has been expanded to 18 core competencies across the domains of data concepts and culture, reading, writing, and comprehension. Databilities® is recognised as the most comprehensive assessment tool of individual data literacy in the world.[1]

The competencies within each domain of the expanded Databilities® framework are:

Data Concepts and Culture *
- Data Culture *
- Data Ethics *

Reading
- Data Discovery
- Evaluating and Ensuring Quality of Data

Writing
- Data Collection
- Data Management and Organisation
- Data Manipulation
- Data Curation and Reuse
- Metadata Creation and Use
- Data Conversion (Format to Format)
- Data Governance *

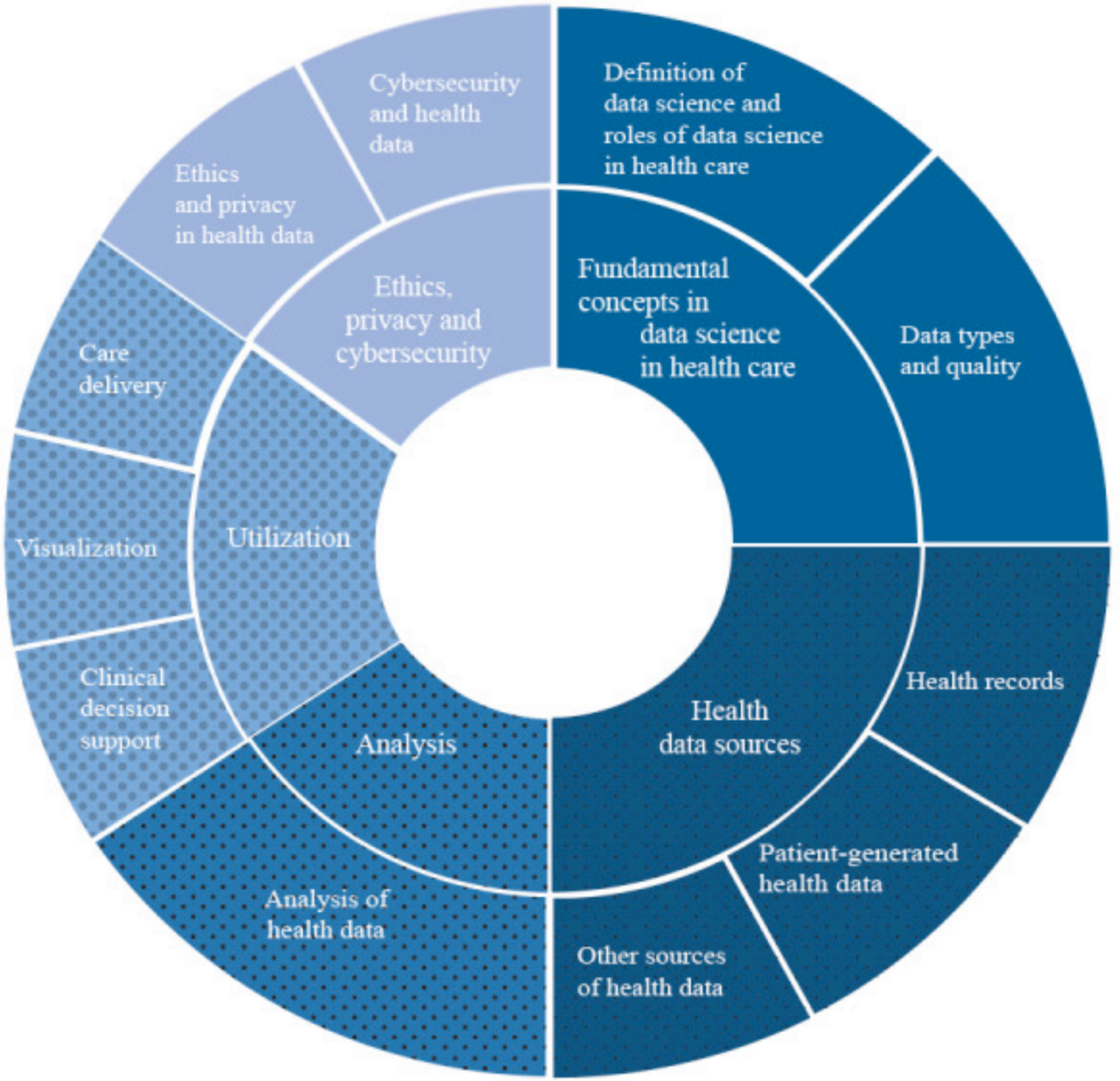
Comprehension
- Data Analysis
- Data Interpretation (Understanding Data)
- Identifying Problems Using Data
- Data Visualisation
- Presenting Data (Verbally)
- Data Driven Decision Making
- Evaluating Decisions / Conclusions Based on Data

For each competency within the Databilities® framework, there are up to 6 levels of progression:

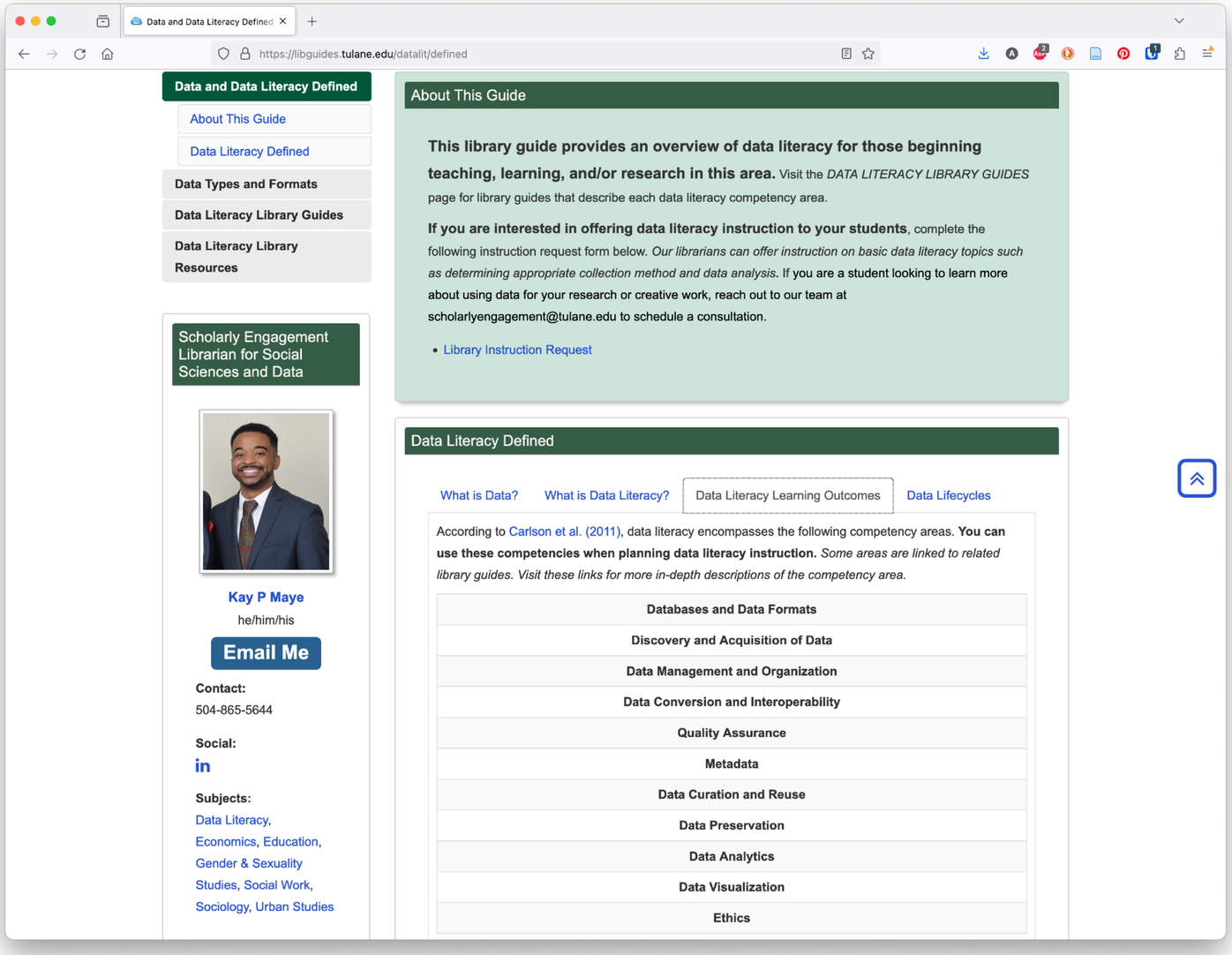
Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
At this level of competency, an individual can complete **simple tasks *with instruction***.	At this level of competency, an individual can complete **simple tasks *on their own, with guidance where needed***.	At this level of competency, an individual can complete **well defined tasks *on their own***.	At this level of competency, an individual can complete **complex problems and tasks *on their own***.	At this level of competency, an individual can ***assist others*** to complete **simple tasks and problems**.	At this level of competency, an individual can ***teach and assist others*** to **complete complex problems and tasks**.

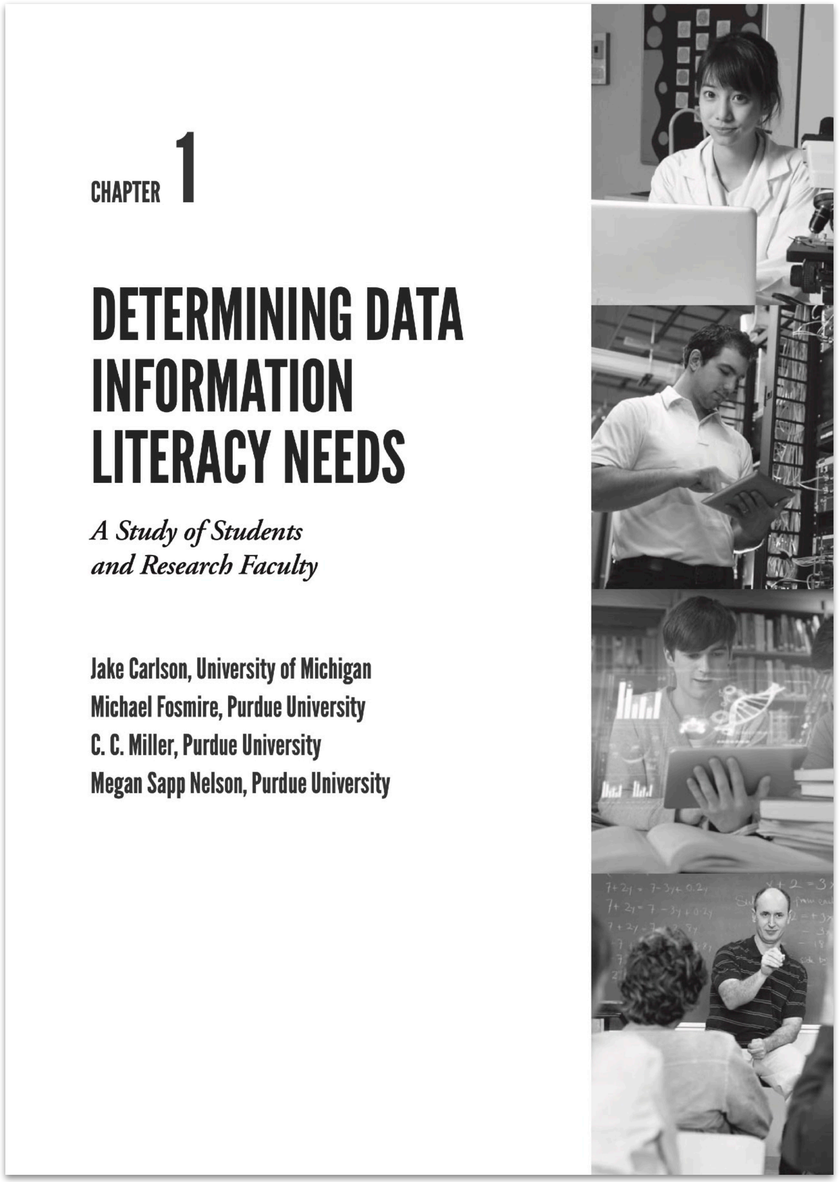
[1] Statistics Canada. 2019. Data Literacy: What It Is and How to Measure It in the Public Service. Statistics Canada Catalogue No. 11-633-X – no. 022. Ottawa. Version updated August 2019.

* Indicates additional competency introduced in expanded Databilities® framework.



Data Science as a Core Competency in Undergraduate Medical Education in the Age of Artificial Intelligence in Health Care. Seth, Hueppchen, Miller, Rudzicz, Ding, Parakh, Record. 2023 https://doi.org/10.2196/46344





Tulane Data Literacy libguide: https://libguides.tulane.edu/datalit/defined
Referencing https://doi.org/10.1353/pla.2011.0022

Data Literacy Competencies for Undergraduate Chemistry

Introduction to Databases

Forming queries in databases to inform research questions or form hypotheses.

Data Discovery

Locating and using data from other researchers, open data sources, or literature.

Data Management

Organizing processed and raw data in separate files. Creating documentation, written procedures or methods for use.

Data Conversion

Changing the structure or format of data files for analysis while protecting against damaging the data.

Quality Assurance

Reviewing data for errors and using consistent protocols and formats during data collection.

Metadata

Understanding the need for adequate description of methods and data and creating documentation to allow others to find, understand and use their data.

Cultures of Practice

Identifying and using data standards and terminologies that are common and accepted across chemistry.

Data Analysis

Knowing and using analysis tools and techniques relevant in chemistry.

Data Visualization

Understanding the accurate use of graphs, plots, diagrams, simulations, and models while avoiding misleading representations.

Ethics of Data

Knowing how to properly cite data to avoid scientific misconduct and data plagiarism.

https://libres.uncg.edu/ir/ecu/listing.aspx?id=32780

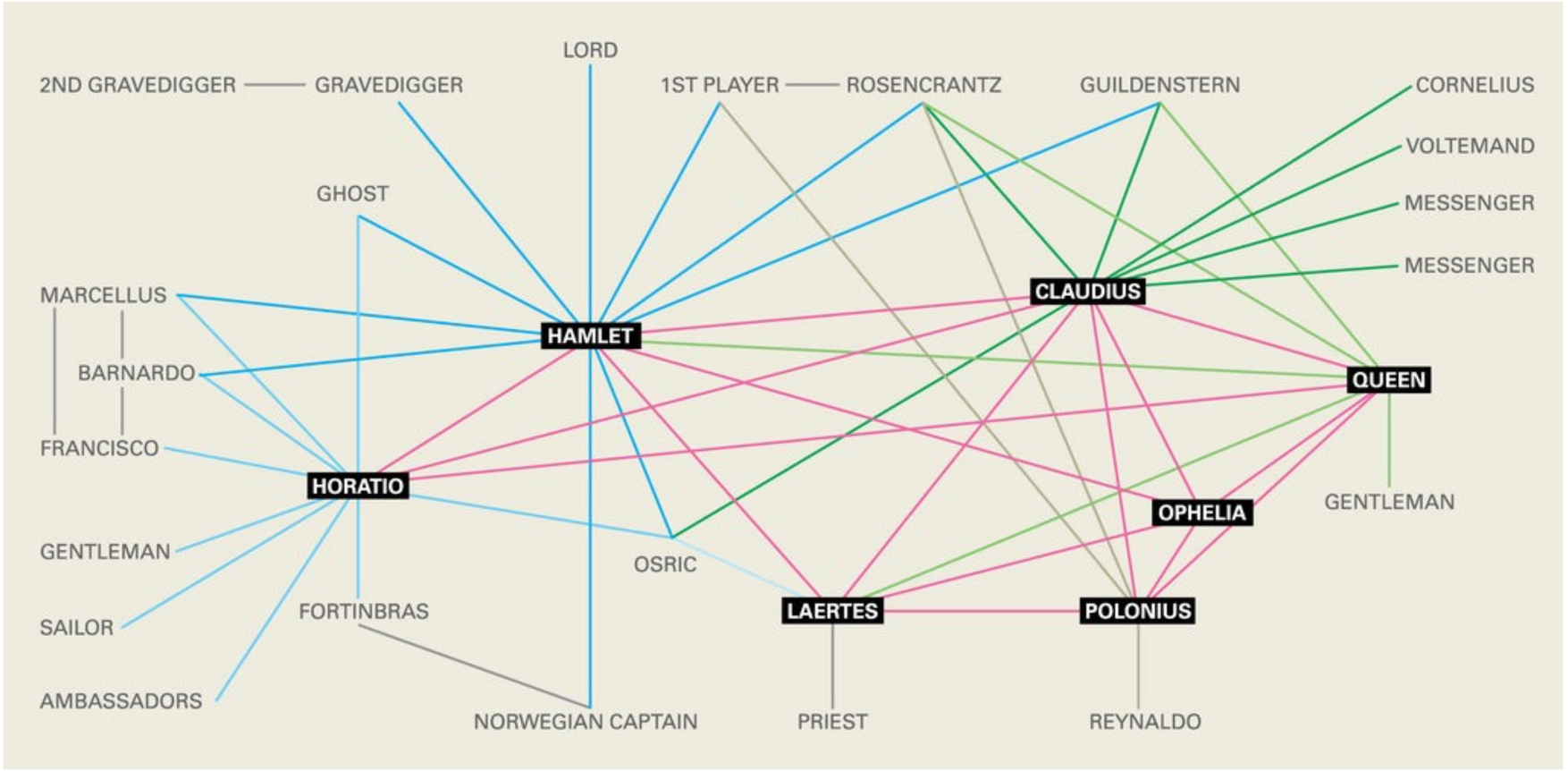
Competencies: we can’t do everything!

- Databases and Data Formats
- Discovery and Acquisition of Data
- Data Management and Organization
- Data Conversion and Interoperability
- Quality Assurance
- Metadata
- Data Curation and Reuse
- Data Preservation
- Data Analytics
- Data Visualization
- Ethics

- We’ll discuss data formatting (particularly tidy data)
- We’ll find some data
-
-
-
-
-
- We’ll do some of this using Excel online
- We’ll try DataWrapper
- We’ll talk about ethics throughout
- We’ll also talk about communicating results

Some motiving examples

What Is Distant Reading?



Kathryn Schulz What is Distant Reading?
https://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html

Inference and Disputed Authorship

The Federalist

Frederick Mosteller and David L. Wallace
With a New Introduction by John Nerbonne

The 1964 publication of Inference and Disputed Authorship made the cover of Time magazine and drew the attention of academics and the public alike for its use of statistical methodology to solve one of American history's most notorious questions: the disputed authorship of The Federalist papers.

Back in print for a new generation of readers, this classic volume applies mathematics, including the once-controversial Bayesian analysis, into the heart of a literary and historical problem by studying frequently-used words in the texts. The reissue of this landmark book will be welcomed by anyone interested in the juncture of history, political science, and authorship.

Frederick Mosteller (1916-2006) was professor of statistics at Harvard University. **David L. Wallace** is professor emeritus of statistics at the University of Chicago.

Contents

12/15/2007

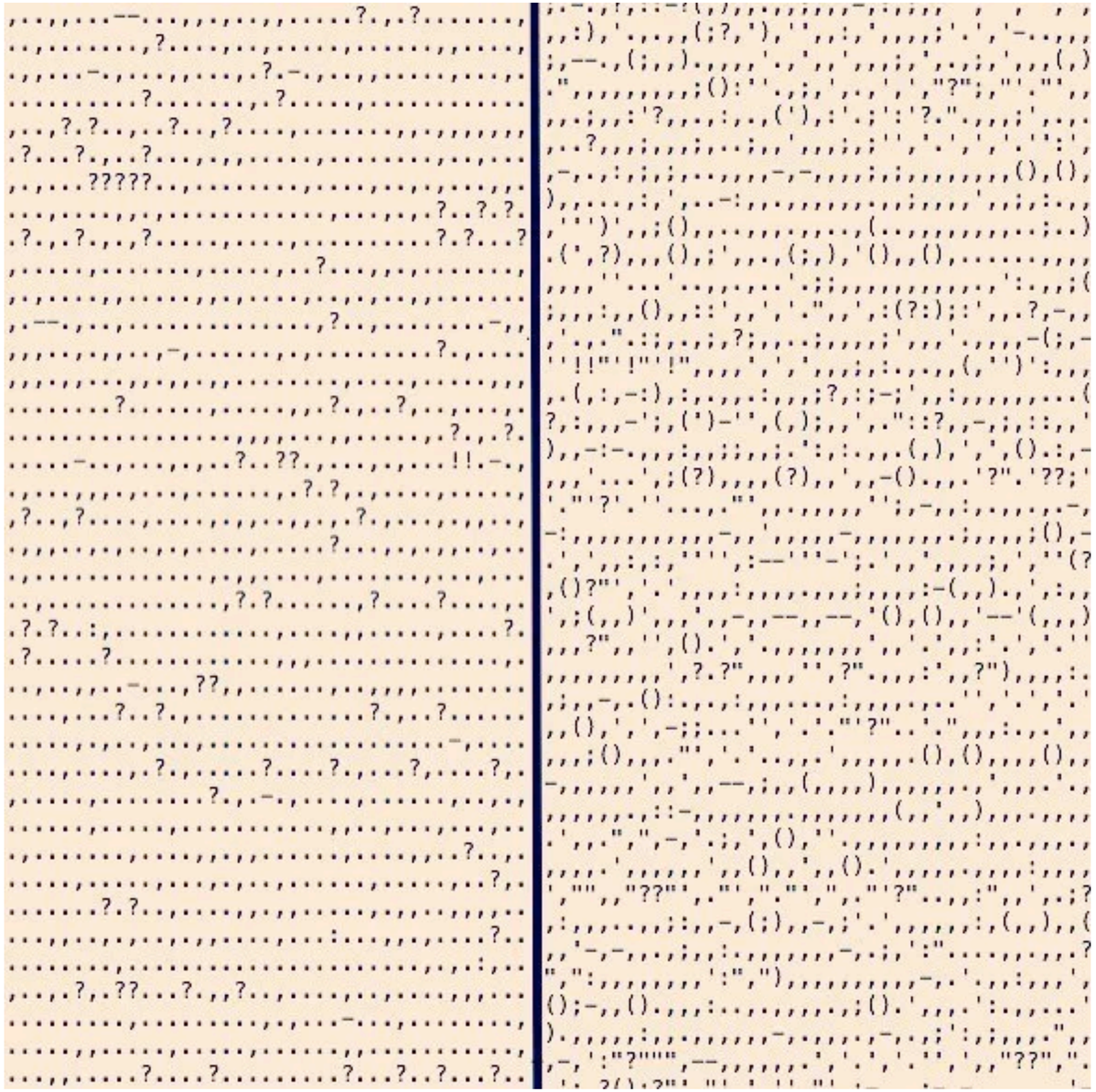
ISBN (Paperback): 1575865521 (9781575865522)
ISBN (electronic): 1575867184 (9781575867182)

Series: The Hume Series
Practical Reflection Latest

pubs @ csli.stanford.edu

CSLI Publications
Stanford University
Cordura Hall
210 Panama Street
Stanford, CA 94305-4101
(650) 723-1839

https://web.stanford.edu/group/cslipublications/cslipublications/site/1575865521.shtml



Punctuation in Blood Meridian by Cormac McCarthy (left) and in Absalom, Absalom! by William Faulkner (right).

Punctuation in novels. Adam J Calhoun

OPINION

A DIARY IN ALPHABETICAL ORDER A TO C

By Sheila Heti

A little more than 10 years ago, I began looking back at the diaries I had kept over the previous decade. I wondered if I'd changed. So I loaded all 500,000 words of my journals into Excel to order the sentences alphabetically. Perhaps this would help me identify patterns and repetitions. How many times had I written, "I hate him," for example? With the sentences untethered from narrative, I started to see the self in a new way: as something quite solid, anchored by shockingly few characteristic preoccupations. As I returned to the project over the years, it grew into something more novelistic. I blurred the characters and cut thousands of sentences, to introduce some rhythm and beauty. When The Times asked me for a work of fiction that could be serialized, I thought of these diaries: The self's report on itself is surely a great fiction, and what is a more fundamental mode of serialization than the alphabet? After some editing, here is the result.
This is part 1 of a 10-part series. Sign up to get it in your inbox.

Actually, he doesn't love you. Actually, he doesn't want you. Actually, he is looking around the world for another girl, and because of who he is, he will find her and be with her. Added in about 4,000 words, bringing it to 56,000. After all, one does have to get back to work. After he left, I lay in bed, hung over, and the sun was shining into my room for the day. After that, I

A Diary in Alphabetical Order, Sheila Heti
https://www.nytimes.com/interactive/2022/03/23/opinion/sheila-heti-alphabetized-diary-abc.html

Literature is not data

The implications of literature as resistance to data extend well beyond the mostly irrelevant little preserve of literature and literary analysis. Algorithms are inherently fascistic, because they give the comforting illusion of an alterity to human affairs. “You don’t like this music? The algorithms have worked it out” is not so far from “You don’t like this law? It works objectively.” Algorithms have replaced laws of human nature, the vital distinction being that nobody can read them. They describe human meanings but are meaningless.

Which is why algorithms, exactly like fascism, work perfectly, with a sense of seemingly unstoppable inevitability, right up until the point they don’t. During the Flash Crash of May 6, 2010, the Dow Jones lost nine percent of its value in five minutes. More recently, Knight Capital lost 440 million dollars at a rate of about 10 million dollars a minute due to what it called “a rogue algorithm.” Algorithms cannot, of course, be rogue. But rogue is the term we have invented for algorithms that don’t do what they’re supposed to, which is as much as to say that their creators don’t comprehend what they’re doing. Before that 440 million dollar loss, Knight Capital had used science to identify a functional law of the marketplace. They had engineered an end to the fundamental human condition of risk. They had not, 45 minutes later. As Borges also wrote, “There is no exercise of the intellect which is not, in the final analysis, useless.” This same futility, it should be remembered, haunts mathematical modeling as much as literary contextualization.

Stephen Marche. Literature is not data: Against the digital humanities
https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/

PRODUCT MARCH 28, 2018

Detecting Crisis: An AI Solution

by Ankit Gupta, Senior Data Scientist

DATA SCIENCE TECH AI MACHINE LEARNING SUICIDE PREVENTION

Content warning: This post references words and phrases associated with suicide, in the context of how Crisis Text Line identifies texters at most imminent risk.

Editor's Note: In July 2017, The Cool Calm presented a post on how Crisis Text Line was using machine learning to triage texters by severity. This post follows up on the evolution of that product.

It's mid-evening on December 1, 2017. A post goes viral on Instagram, resulting in record texter volume at Crisis Text Line. Hundreds of volunteer Crisis Counselors pour into the system to respond. An opening message from one texter reads (paraphrased for confidentiality):

"I just took an overdose of Lithium and I'm letting it build up in my system for a few days."

The triaging algorithm flags this message as high-risk for a suicide attempt, and moves it straight to the top of the queue. A Crisis Counselor responds within 20 seconds. Within an hour, the texter has been located by emergency services, and is safe. This is the power of data at scale. Here's how we did it:

https://www.crisistextline.org/blog/2018/03/28/detecting-crisis-an-ai-solution/

Mental Health Mental Health

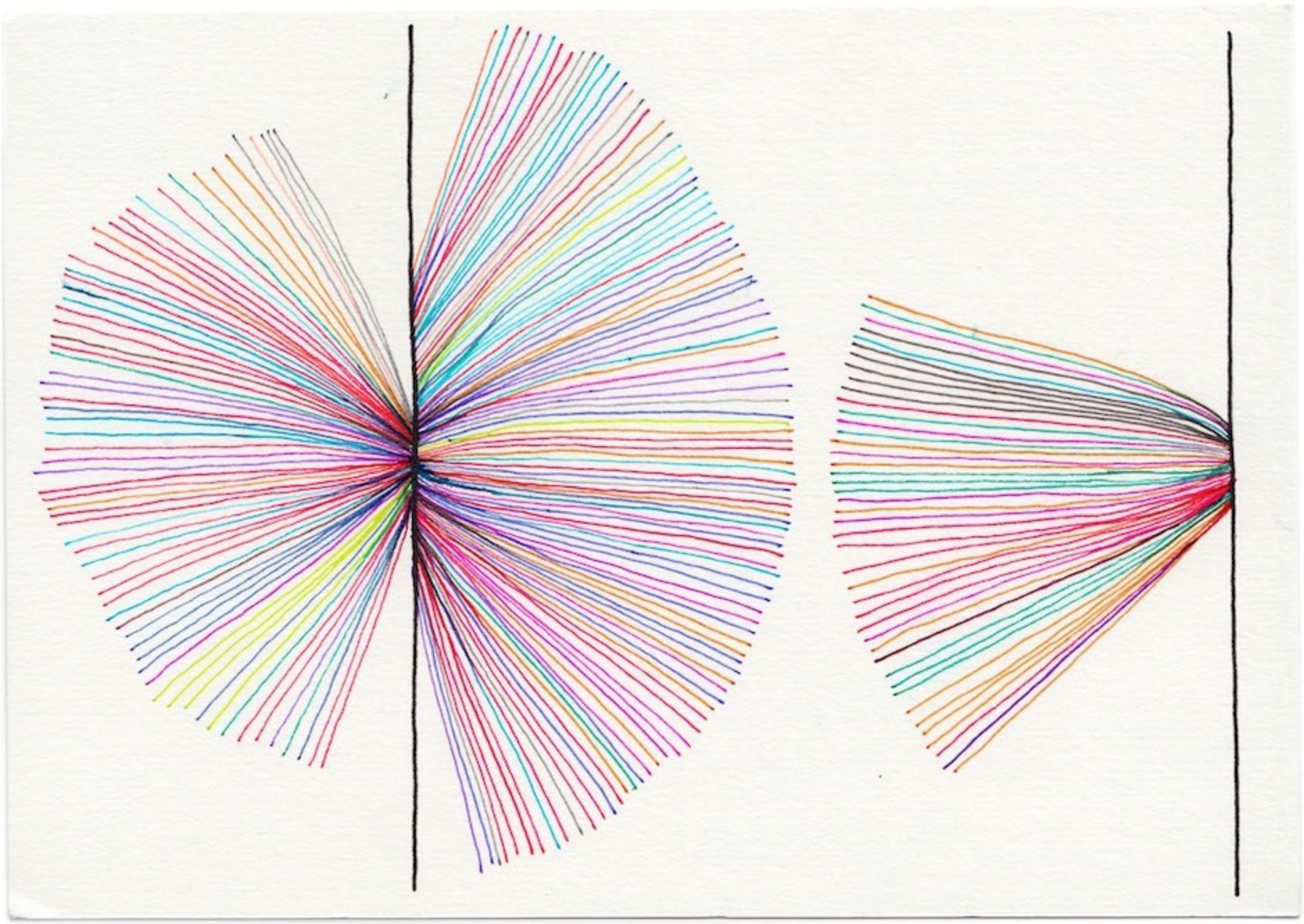
Crisis Text Line tried to monetize its users. Can big data ever be ethical?

The crisis intervention service had concerns about its financial future, but made a huge mistake.

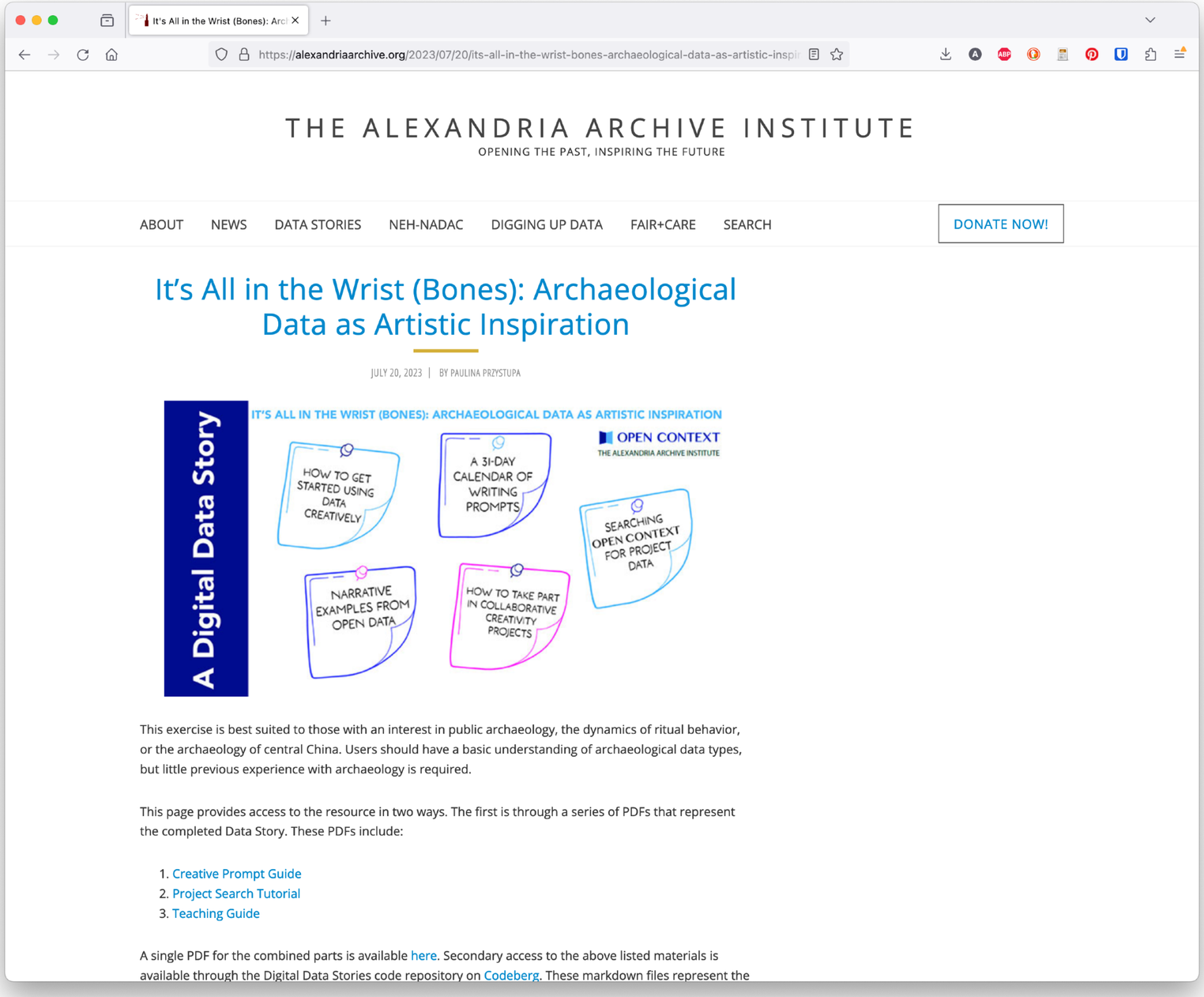
By Rebecca Ruiz on February 3, 2022

Crisis Text Line tried to make a business out of user data. It went terribly wrong. Credit: Vicky Leta / Mashable

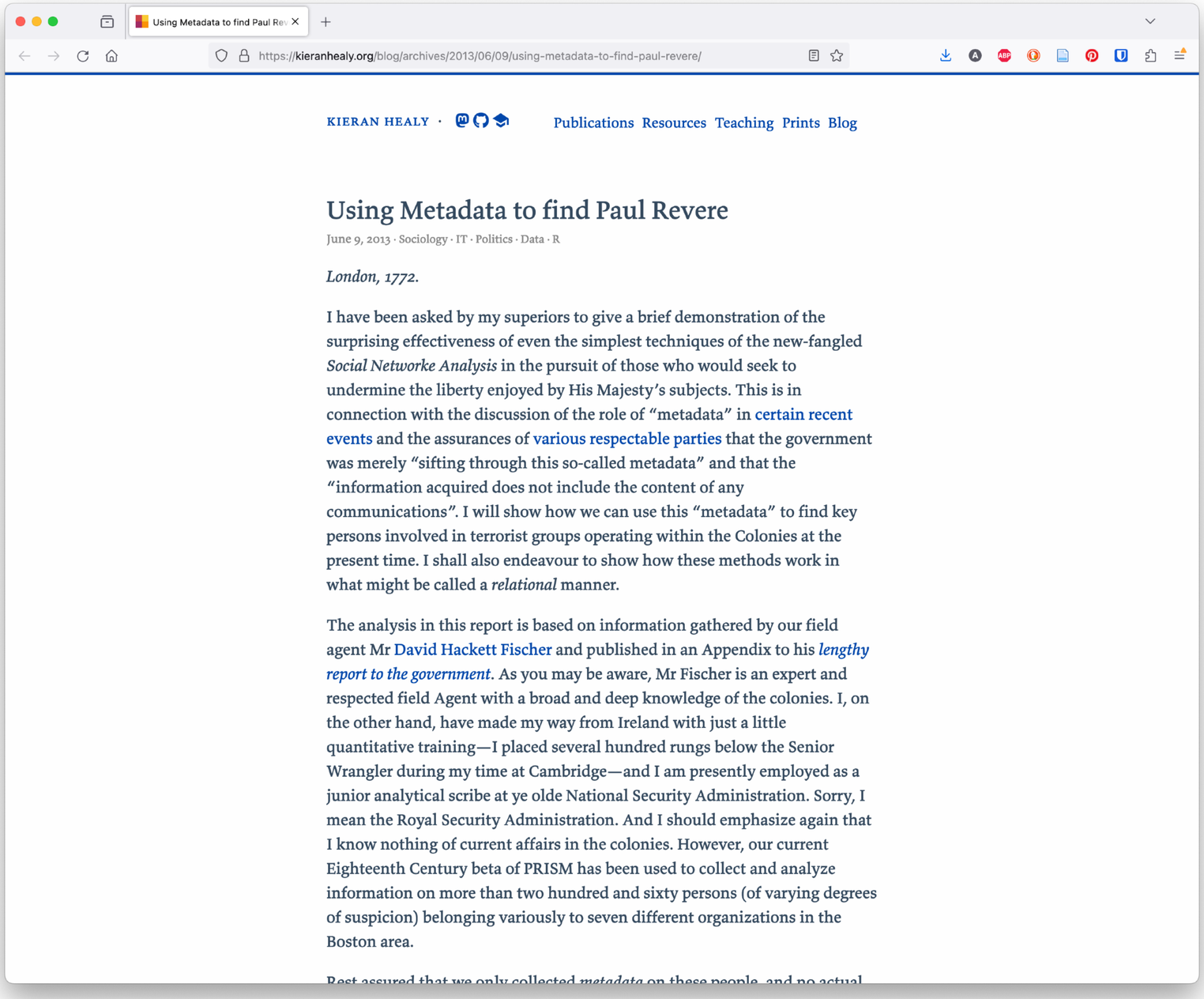
https://mashable.com/article/crisis-text-line-loris-big-data-ethics



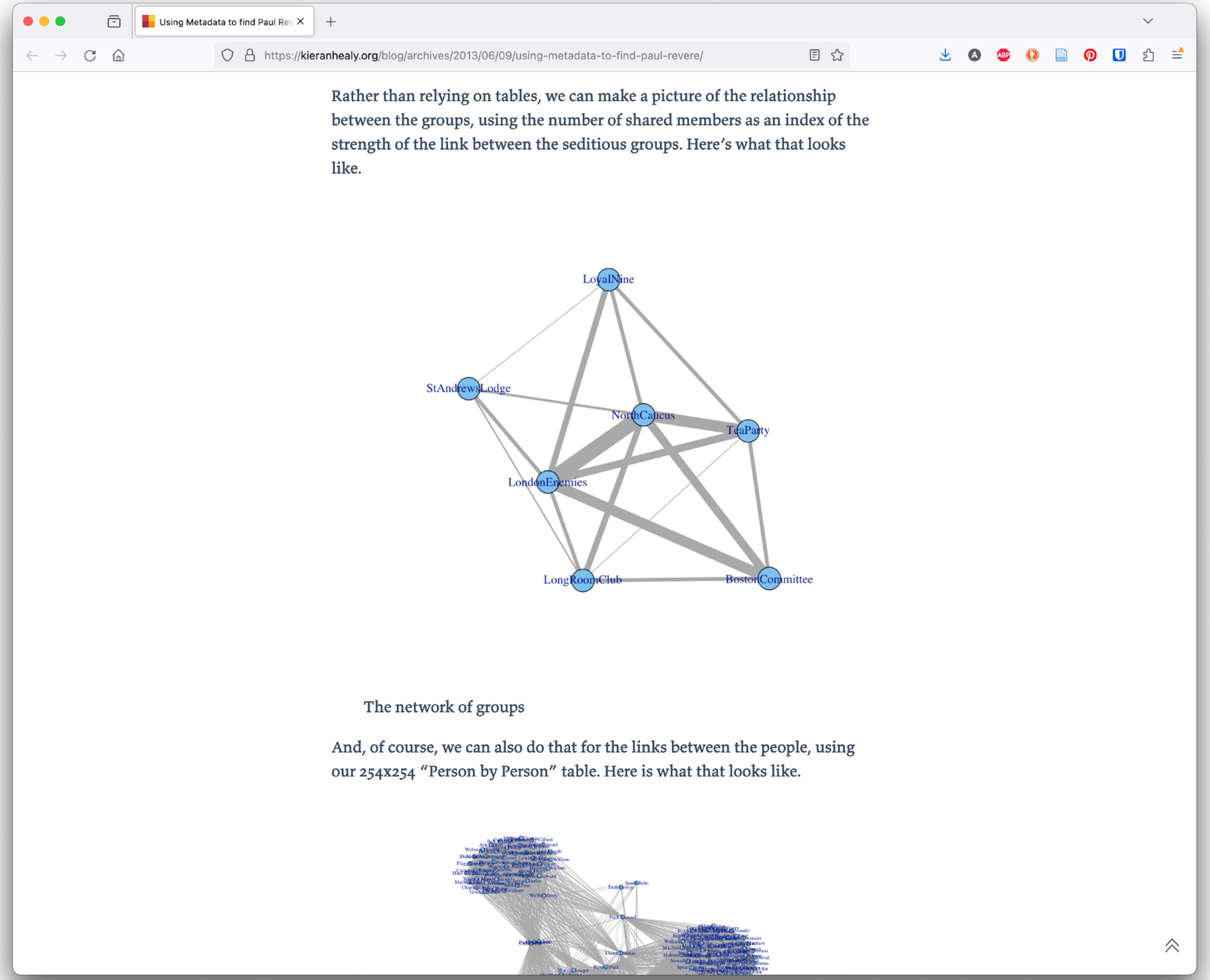
Giorgia Lupi and Stefanie Posavec, Dear Data www.dear-data.com



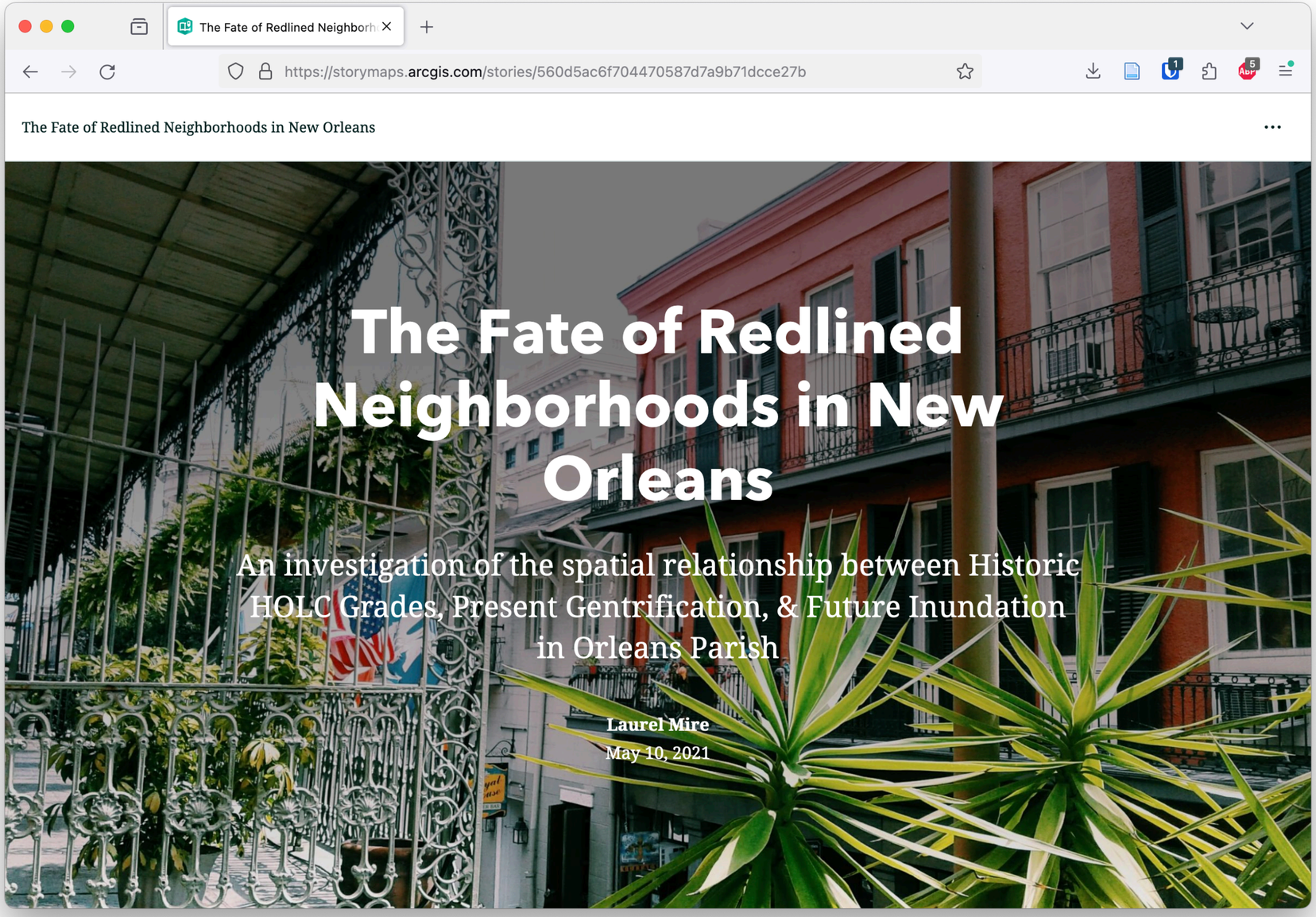
It's All in the Wrist (Bones): Archaeological Data as Artistic Inspiration. Paulina Przystupa. https://alexandriaarchive.org/2023/07/20/its-all-in-the-wrist-bones-archaeological-data-as-artistic-inspiration/



Using Metadata to find Paul Revere. Kieran Healy.
https://kieranhealy.org/blog/archives/2013/06/09/using-metadata-to-find-paul-revere/



Using Metadata to find Paul Revere. Kieran Healy.
https://kieranhealy.org/blog/archives/2013/06/09/using-metadata-to-find-paul-revere/



The Fate of Redlined Neighborhoods in New Orleans. Laurel Mire.
https://storymaps.arcgis.com/stories/560d5ac6f704470587d7a9b71dcce27b

Jer Thorpe, St. Louis Map Room https://blprnt.medium.com/making-mapping-more-human-77a96e92ed49

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 63, No. 6, June 2017, pp. 1657–1672
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Reference-Dependent Preferences: Evidence from Marathon Runners

Eric J. Allen,[a] Patricia M. Dechow,[b] Devin G. Pope,[c] George Wu[c]

[a] Marshall School of Business, University of Southern California, Los Angeles, California 90089; [b] Haas School of Business, University of California, Berkeley, Berkeley, California 94705; [c] Booth School of Business, University of Chicago, Chicago, Illinois 60637

Contact: eric.allen@marshall.usc.edu (EJA); patricia_dechow@haas.berkeley.edu (PMD); dpope@chicagobooth.edu (DGP); wu@chicagobooth.edu (GW)

Received: March 30, 2015
Revised: August 17, 2015
Accepted: October 27, 2015
Published Online in Articles in Advance: April 20, 2016

https://doi.org/10.1287/mnsc.2015.2417

Abstract. Theories of reference-dependent preferences propose that individuals evaluate outcomes as gains or losses relative to a neutral reference point. We test for reference dependence in a large data set of marathon finishing times ($n = 9{,}789{,}093$). Models of reference-dependent preferences such as prospect theory predict bunching of finishing times at reference points. We provide visual and statistical evidence that round numbers (e.g., a four-hour marathon) serve as reference points in this environment and as a result produce significant bunching of performance at these round numbers. Bunching is driven by planning and adjustments in effort provision near the finish line and cannot be explained by explicit rewards (e.g., qualifying for the Boston Marathon), peer effects, or institutional features (e.g., pacesetters).

History: Accepted by John List, behavioral economics.
Funding: The authors thank the John Templeton Foundation (New Paths to Purpose project) for generous financial support.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mnsc.2015.2417.

Keywords: reference dependence • prospect theory • loss aversion • bunching • effort provision

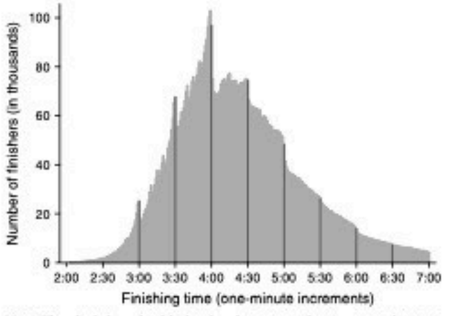
1. Introduction

Recent theories of economic behavior propose that the evaluation of an outcome may be affected by comparisons of that outcome with a reference point and not merely tastes, risk attitudes, and wealth levels, as in classical economic models. For example, how an employee views a bonus of $1,000 might depend on the level of previous bonuses, what bonuses were distributed to other members of the organization, or the employee's expectations about what bonuses were possible (Card et al. 2012, Kahneman 1992, Kőszegi and Rabin 2006).

A reference point divides outcomes into gains or losses, thus creating a qualitative difference in the valuation of outcomes slightly above or below that reference point. We suggest that the distinguishing feature of reference-dependent models is some form of discontinuity at the reference point that is psychologically based and not the result of an extrinsic benefit. For example, a primary feature of prospect theory, the most well-known and influential account of reference-dependent preferences, is loss aversion (Kahneman and Tversky 1979, Tversky and Kahneman 1992). The premise that "losses loom larger than gains" (Kahneman and Tversky 1979) has implications for a wide range of economic activities, including risky decision making, choice of consumption bundles, and effort provision (DellaVigna 2009, Tversky and Kahneman 1991). A second property, diminishing sensitivity, is captured by prospect theory's characteristic S-shaped value function that is concave for gains and convex for losses. Although prospect theory is the most prominent model of reference dependence, the discontinuity at the reference point in some instances might instead be produced by a jump (or "notch") in the utility function at the reference point.

Researchers have moved beyond Kahneman and Tversky's laboratory demonstrations of reference dependence to explain behavioral anomalies across a wide variety of field settings.[1] In a recent review of prospect theory, Barberis (2013) highlighted the key challenge to researchers testing for field evidence of reference-dependent preferences: it is often difficult to know exactly what reference points are relevant for individuals in field settings. The difficulty in identifying the appropriate reference point is best illustrated by a stream of work examining the possible role that reference points play in labor supply and effort provision. Camerer et al. (1997) argued that taxi drivers have a downward-sloping labor supply curve induced by daily income targets (see also Fehr and Goette 2007, and Mas 2006). This paper led to additional analyses that used different data sets and econo-

Figure 2. Distribution of Marathon Finishing Times ($n = 9{,}789{,}093$)



Note. The dark bars highlight the density in the 1-minute bin just before each 30-minute threshold.

4:00 marks, compared to 74,968, 69,648, and 67,861 finishers in the 4:00, 4:01, and 4:02 bins. Although the four-hour mark is particularly dramatic, qualitatively similar differences exist at other hour and half-hour marks and, to a lesser extent, at 10- and 15-minute marks. There are 50.0%, 21.5%, and 29.5%, more finishers in the 1-minute bin before 3:00, 3:30, and 4:00, respectively, than the 1-minute bin after these round numbers. This excess mass measure for 10-minute marks is less dramatic but still substantial: 11.9%, 8.5%, 9.4%, and 7.0% for 3:10, 3:20, 3:40, and 3:50, respectively.[11]

We next measure the amount of excess mass around the round number reference point and test whether the excess mass is statistically significant by adapting a methodology proposed in Chetty et al. (2011) to quantify the extent of excess mass in an interval around a round number.[12] We draw an analogy between our setting and individual taxpayer responses to "kinks" in the tax code (e.g., Kleven and Waseem 2013, Saez 2010). Consistent with the hypothesis that income will bunch around tax rate thresholds, Chetty et al. (2011) found that Danish taxpayers bunch around the income cutoff for the top marginal income tax rate. In our setting, we hypothesize that round number reference points serve as a discontinuity in a marathoner's utility function in a similar manner to how income thresholds do for taxpayers (see Section 2). As in Chetty et al. (2011), the observed bunching is likely to be diffuse rather than a point mass. Runners are unable to perfectly control their effort levels over the course of the race. They may underestimate the amount of energy they have left, incorrectly calculate the required pace to meet the benchmark, or build a cushion into their pacing that causes them to beat the reference point by more than a small amount. As a result, rather than seeing a sharp increase in runners just beating the reference point and then an immediate drop (as required by Propositions 2.1 and 2.2), we expect to see somewhat diffuse bunching of finishing times around the reference point. This dispersion will reflect runners who attempt to meet the reference point and just miss, as well as those who beat it by a few minutes.

To calculate the amount of bunching, we follow the Chetty et al. (2011) methodology. The counterfactual distribution is estimated by fitting a quintic polynomial to the local density of finishing times around the reference point excluding the bunching region. The difference between the actual density in the bunching region and the fitted counterfactual density is the excess number of finishers around the reference point, with the standard error for the amount of excess mass determined by a bootstrap procedure. Throughout, we take the local window around each potential round-number reference point to be 16 minutes (8 minutes before a round number and 8 minutes after a round number). For example, to test for bunching at 3 hours and 30 minutes, we use a window from 3 hours and 22 minutes to 3 hours and 38 minutes. We choose this window to avoid bunching that may occur at a 10-minute mark in the counterfactual distribution either above or below the reference point of interest. We look for evidence of bunching itself in a 4-minute window right before each round number. As recommended by Chetty et al. (2011), this window was chosen based on visual inspection of the bunching. We employ a conservative test and use the same window for every potential reference point. Finally, before calculating the excess mass measure, we shift the entire counterfactual distribution upward so that the area underneath the counterfactual curve is equivalent to the area under the actual density function, thus avoiding the bias that would otherwise occur since the bunching is likely drawing from individuals just outside the bunching region. Thus, without this correction, we would essentially be double counting runners that are bunching at the reference point and causing the counterfactual distribution to be lower than it would otherwise be in a truly counterfactual world.

The main results of the bunching estimation applied to our full sample are depicted in Figure 3 and summarized in Table 2. Figure 3 graphically shows the 16-minute window around reference points at 3:00, 3:10, 3:20, 3:30, 4:00, 4:30, 5:00, and 6:00. The actual finishing times are plotted in 15-second intervals along with the counterfactual distribution that we estimate using the procedure above. The figures show clear evidence of bunching at the majority of the round number reference points. The bunching is particularly evident at the 3- and 4-hour marks. In Sections A.5–A.7 of the online appendix, we present the same results for all 10- and 15-minute marks from 2:30 to 6:00 and show that our results are robust to variations in the

"Liberal arts module" from SDS 192: Introduction to Data Science. Collaboration between Ben Baumer (Statistical and Data Sciences) and Alex Keller (Film Studies)

Mini-Project

You will investigate a Film and Media Studies question of interest. Your deliverable will be a 5-minute class presentation (to your peers) and a blog post written in R Markdown.

In groups of three, please choose a topic from the following list illuminated by Prof. Keller:

1. Westerns over time: The Western is a central U.S. genre, not just to Hollywood and mainstream film (think *Stagecoach*, 1939 and *The Searchers*, 1956), but also to independent film production (*Brokeback Mountain*, 2005). Film critics have pronounced the Western dead repeatedly, and, so far, always wrongly. Between 1900 and 2015, when and how has Western film production peaked and ebbed? What might account for this, and does the IMDB define a Western in the 1950 in the same way as it does in 2000?
    - See `info_type_id = 3 AND info = 'Western'`
2. Genre Multiplicity: Prof. Keller explained how the labor model in Hollywood has changed over time, specifically with respect to the 1947 court decision that broke up the vertical integration of studios. Is it true that movies made today tend to have overlapping genres in a way that they did not in the past? Is is true that older movies tended to fit squarely in one genre, while more recent movies tend to span multiple genres? Is there a historical moment where this genre complexity begins? Does it have patterns of any discernible kind?
    - See `info_type_id = 3`
3. Sequels with SQL: Hollywood seems to be more repetitive than it used to be. Is this true? We now have not only sequels, but also remakes and reboots. Are these follow-up films more common today than they used to be? Can you trace the evolution of sequels across time in Hollywood? Are sequels more frequent today? Is the spacing between a sequel and its original shorter than it used to be?
    - See `movie_link` table
4. Community Detection and Missingness: In statistics, data are often missing. Yet the difference between data that are *missing at random* versus *missing systematically* is crucial. What movies are present in the IMDB? We know there are box office Hollywood movies, independent films, experimental/avant-garde films, adult films, and foreign films. What films are missing? How does a film merit inclusion in the imdb.com data base? Can you think of any films that are not in the database? Are there entire areas of moving image production that imdb.com simply doesn't recognize (e.g. youtube.com)? If so, what accounts for that? Could you create a visualization showing the source of every film in the IMDB? What communities are present? What communities are not present? When you "map" these data, look beyond geography.

1E. Archival Undercurrents

> Alexis Easley & Cecilia Becicka, University of St. Thomas, "Victorian Authorship, Geography, and Gender: Insights from the Chambers's Archive"

The W. & R. Chambers archive at the National Library of Scotland includes two ledgers that list the names, addresses, article titles, and remuneration for the 164 women writers who contributed to the journal between 1839 and 1855. In this presentation, we will share insights gleaned from analysis of this data – e.g., where these women contributors lived, how much they were paid, and what sort of writing they published in the journal. We also conduct a comparative analysis of male and female contributors with special attention to the genres of writing they contributed and the remuneration they received. By mapping the addresses of women contributors, we identify their social and literary networks in Dublin, Cork, London, and Edinburgh. We also show how their correspondence addresses change over time. The Chambers archive also includes correspondence from many of the women contributors whose names appear in the ledger. In the second part of this presentation, we analyze their interactions with the Chambers firm – from submitting their work to negotiating payment, working with editors, and adapting to specific editorial requirements such as article length. We reveal some intriguing stories behind these anonymous publications – e.g., how Agnes Loudon (daughter to the famous botanist) published her first story in the journal at age thirteen and how Janet Wills used publication in the journal to spoof her own experience as the wife of a newspaper editor (W.H. Wills).