Sophia Toubian

IS 298C

June 14, 2024

The move towards "Data" and Claude Shannon

I. Research Problem

The history of "data" as a universal term for information is a fascinating journey marked by evolving perceptions and practices surrounding information categorization and theorizing. This research explores the historical evolution of information labeling, and the emergence of "data" as the predominant term for all kinds of information. It focuses on Claude Shannon, who shaped the discourse around data, spanning developments in statistics, computer science, information theory, and societal attitudes. The research aims to uncover the sociocultural, theoretical, and technological forces driving the shift from diverse terminology to the universal adoption of "data" - a transition that reflects broader changes in how information is conceptualized and utilized, with profound impacts across science, technology, and the humanities.

The key questions guiding this research are:

1. How did the conceptualization and labeling of information evolve over time, leading to "data" becoming dominant?

2. What were the main technological, societal, and intellectual factors influencing this transition?

3. How did specific disciplines and industries contribute to shaping the data discourse?

4. What implications did the universalization of "data" have for knowledge production, governance, and the economy?

Answering these questions is crucial for enriching the field of information studies with a more nuanced understanding of data's conceptual origins and historical significance. It challenges assumptions of data as a neutral, objective concept, instead highlighting the complex human processes underlying its construction and interpretation.

II.   Literature Review

This research draws upon a diverse range of works from the history and philosophy of information, data studies, and related fields to examine the complex genealogy and implications of data. The literature can be organized into several key themes: the conceptual boundaries of data and documents, the social construction of data and scientific knowledge, the rise of big data and its disciplinary implications, the circulation of data across contexts, the historical foundations of information theory, and the rhetorical and political dimensions of data.

Conceptual Boundaries of Data and Documents

Michael Buckland's article "What is a 'Document'?" (1997) and Maurizio Ferraris' book "Documentality: Why It Is Necessary to Leave Traces" (2012) offer valuable perspectives on the conceptual boundaries of documents and data. Buckland argues for an expanded definition of documents that includes objects and events, opening up possibilities for broader understandings of information and data. Ferraris introduces the concept of "documentality," emphasizing the ontological and social dimensions of documents as constitutive of reality itself. These works provide a foundation for considering the nature and scope of data as a form of documentation that actively shapes the world.

Social Construction of Data and Scientific Knowledge

The social construction of data and scientific knowledge is a central theme in the literature. Bruno Latour and Steve Woolgar's "Laboratory Life" (1979) offers an ethnographic

study of scientific practice, revealing how facts and data emerge through social interactions and material practices within the laboratory. Ian Hacking's "The Social Construction of What?" (1999) further elaborates on the philosophical underpinnings of social constructionism, examining how classifications and categories structure knowledge. These works underscore the situated and contingent nature of data production, challenging notions of scientific objectivity and neutrality.

<u>Rise of Big Data and Its Disciplinary Implications</u>

The rise of big data and its implications across disciplines is another significant theme. Jeremy Hunsinger's chapter "Big Capta?" (2019) stresses the importance of critically examining knowledge construction through data collection and analysis, employing the term "capta" to emphasize the need for interpretive frameworks beyond mere data points. Sabina Leonelli's *Data-Centric Biology: A Philosophical Study* (2016) provides an empirical case study of the increasing data-centricity in the life sciences, highlighting the complex interplay between data, theory, and practice in scientific knowledge production.

<u>Circulation of Data Across Contexts</u>

The circulation of data and scientific ideas across boundaries is a key concern in the literature. James Secord's "Knowledge in Transit" (2004) offers a framework for understanding how the meaning and significance of scientific ideas are transformed as they travel across diverse audiences and settings. Bruce Lewenstein's "From Fax to Facts: Communication in the Cold Fusion Saga" (1995) provides a case study of how scientific communication shapes public understanding of controversial claims, demonstrating the role of media coverage and informal communication networks in mediating the circulation of data and evidence.

<u>Historical Foundations of Information Theory</u>

The pioneering work of Claude Shannon in information theory is central to understanding the historical foundations of data. Shannon's "A Mathematical Theory of Communication" (1948) and "Prediction and Entropy of Printed English" (1951) introduced key concepts like entropy, noise, and channel capacity, laying the groundwork for later developments in data science and communication engineering. Chris Wiggins and Matthew Jones' *How Data Happened: A History from the Age of Reason to the Age of Algorithms* (2020) offers a comprehensive history of how "data" became a ubiquitous term, unpacking the sociocultural, technological, and intellectual forces driving this linguistic and conceptual shift.

Rhetorical and Political Dimensions of Data

The rhetorical and political dimensions of data are explored in several works. Johanna Drucker's "Humanities Approaches to Graphical Display" (2011) introduces a critical lens to data visualization, emphasizing the role of subjective interpretation and advocating for a humanistic approach that foregrounds ambiguity and rhetoric. Lauren Klein and Miriam Posner's essay "Data" (2020) traces the early uses of "data" in fields like theology and mathematics, highlighting its rhetorical function as a means of establishing ground truth. They also discuss the ways in which data has been wielded by marginalized communities as a tool for advocacy and resistance, while cautioning against its reductive and dehumanizing potential.

By organizing the literature thematically, we can see how these works collectively contribute to a nuanced understanding of data's complex nature and implications. The conceptual boundaries of data and documents are explored, revealing their active role in shaping reality. The social construction of data and scientific knowledge is emphasized, challenging notions of objectivity and neutrality. The rise of big data and its disciplinary implications are examined,

highlighting the need for critical interpretive frameworks. The circulation of data across contexts is investigated, tracing the transformation of meaning as ideas travel. The historical foundations of information theory are established, providing a basis for understanding the development of data science. Finally, the rhetorical and political dimensions of data are foregrounded, revealing its potential for both oppression and empowerment.

Engaging with these diverse perspectives allows for a more comprehensive and historically-informed understanding of data's genealogy and its far-reaching consequences for knowledge production, power relations, and social justice. By drawing upon insights from the history and philosophy of information, data studies, and related fields, this research aims to contribute to ongoing debates and discussions about the nature, use, and implications of data in contemporary society.

III.   Primary Source Materials

The extensive archives of Claude Shannon's work, housed primarily at the Library of Congress and supplemented by collections at the University of Southern California, serve as the main primary source for this research. As a pivotal figure in 20th-century information science, Shannon made groundbreaking contributions to communication theory, cryptography, and early computing during his tenure at Bell Labs and beyond. His wide-ranging archive, which includes correspondence, research notes, manuscripts, lectures, patents, photographs, and even personal artifacts like juggling props and chess machines, offers a rich and multifaceted resource for investigating the historical evolution of data as a concept and its broader implications for various fields.

For the purposes of this research, several key materials from Shannon's archive will be examined to shed light on the research questions at hand. Firstly, Shannon's seminal scientific

publications, such as "A Mathematical Theory of Communication" (1948) and "Prediction and Entropy of Printed English" (1951), will be closely analyzed. These works articulate Shannon's pioneering ideas on information theory, introducing concepts like entropy, noise, and channel capacity that laid the conceptual foundations for data science. By engaging with these texts, we can trace the emergence of a quantitative and statistical approach to information that would later be adapted and expanded in the field of data studies.

In addition to his published works, Shannon's correspondence with key contemporaries like Alan Turing, John von Neumann, Norbert Wiener, and Warren Weaver will be examined to reveal the intellectual influences and collaborations that shaped his theories. These exchanges offer valuable insights into the broader scientific and philosophical context in which Shannon's ideas developed, highlighting the interdisciplinary nature of his work and its engagement with cutting-edge research in mathematics, engineering, and computer science. By situating Shannon's contributions within this larger intellectual milieu, we can better understand the historical forces and conceptual shifts that gave rise to the modern notion of data.

Shannon's lectures and manuscripts aimed at explaining his work to broader audiences will also be consulted to shed light on how he translated complex mathematical and technical concepts for non-specialist readers. These materials provide a window into Shannon's own understanding of the wider implications and potential applications of his theories across various fields, from communication engineering to linguistics and beyond. By examining how Shannon framed his ideas for different audiences, we can gain insight into the processes by which his work was disseminated and adapted beyond its original disciplinary context.

Alongside these more polished presentations of his work, Shannon's notes, drafts, and working papers will be carefully studied to trace the evolution of his thinking and the emergence

of key terms and frameworks that would later be taken up in data studies. These materials offer a glimpse into the iterative and often messy process of scientific discovery, revealing the false starts, revisions, and conceptual breakthroughs that led to Shannon's most influential ideas. By attending to the genealogy of Shannon's thought as preserved in these informal writings, we can develop a more nuanced understanding of the historical and intellectual context in which the modern concept of data took shape.

Finally, Shannon's patents and technical reports will be examined to explore the real-world implementations and practical applications of his theoretical insights in the domains of communication engineering and early computing. These materials showcase how Shannon's ideas were translated into concrete technologies and systems, laying the groundwork for the widespread digitization of information that would transform multiple industries in the decades following his groundbreaking work. By tracing the material instantiations of Shannon's theories, we can better understand the far-reaching impact of his work and its role in shaping the technological infrastructure that underlies the data-driven world we inhabit today.

To fully leverage the richness and depth of Shannon's archive, comprehensive access to the collections held at the Library of Congress and the University of Southern California is essential. Visiting these sites and working closely with archivists and subject specialists will be necessary to ensure thorough coverage of relevant materials and to uncover potentially overlooked sources that can further illuminate Shannon's intellectual trajectory and its broader significance. The expertise of these professionals, along with their intimate knowledge of the collections, will be invaluable in identifying key documents and navigating the vast array of materials available.

By engaging with this diverse range of primary sources from Shannon's archive, this research aims to shed new light on the historical evolution of data as a concept and its far-reaching implications for fields ranging from information science and computer engineering to the humanities and social sciences. Through close reading and contextual analysis of Shannon's published works, correspondence, lectures, drafts, and technical materials, we can trace the emergence of a quantitative and statistical approach to information that would later be formalized as data science. At the same time, by situating Shannon's contributions within the broader intellectual and social context of his time, we can develop a more nuanced understanding of the complex interplay of ideas, technologies, and institutional forces that shaped the modern data landscape and laid the groundwork for the modern pivot to "data.".

Ultimately, by grounding our investigation in the rich primary materials preserved in Shannon's archive, this research aims to contribute to ongoing debates and discussions about the nature, use, and implications of data in contemporary society. By excavating the historical roots of our current data-centric paradigm and attending to the contingent and context-dependent nature of Shannon's foundational ideas, we can develop a more critical and reflexive approach to data that recognizes its rhetorical dimensions, its entanglement with structures of power and inequality, and its potential for both reifying and resisting dominant narratives and ways of knowing. In doing so, this research seeks to contribute to the growing field of critical data studies and to enrich our understanding of the complex genealogy of data as a concept and a material practice.

IV.   Research Methods and Interpretive Framework

This research employs a combination of historical methods and interpretive frameworks drawn from the history and philosophy of science, science and technology studies (STS), and

critical data studies to investigate the intellectual history of data as a concept, focusing on the work and archives of Claude Shannon. By engaging with primary sources from Shannon's archive, including his published works, correspondence, lectures, drafts, and technical materials, this research aims to trace the emergence and evolution of data as a theoretical construct and a material practice, situating it within the broader social, institutional, and technological contexts of mid-20th century information science and communication engineering.

Central to this research is a methodological commitment to examining the social construction of scientific knowledge, which understands data not as an objective or given entity, but rather as the product of specific historical contexts, communities, and practices. Drawing on the work of theorists like Bruno Latour and Simon Schaffer, this research attends to the ways in which scientific facts and concepts emerge through complex networks of social interactions, material practices, and institutional arrangements. By applying this framework to the case of Shannon's work on information theory, we can interrogate the processes by which data was conceptualized, quantified, and operationalized within the specific intellectual and institutional milieu of Bell Labs and the larger field of communication engineering.

To further situate Shannon's contributions within this broader context, this research employs biographical methods to examine the role of key individuals in shaping scientific fields and theoretical paradigms. By closely examining Shannon's unique interdisciplinary background, which spanned mathematics, electrical engineering, and cryptography, as well as his collaborations with prominent figures, we can better understand the intellectual influences and social networks that informed his groundbreaking work on information theory. Drawing on the approaches of scientific biographers like Walter Isaacson and Ray Monk, this research attends to

the ways in which Shannon's personal and professional trajectory shaped his conceptual innovations and their subsequent impact on the field of information science.

In addition to examining the social construction of scientific knowledge and the role of individual actors, this research also traces the circulation of scientific ideas and concepts across disciplinary boundaries and into the public sphere. Using James Secord's model of "knowledge in transit," which understands the movement of scientific ideas as a complex process of translation, appropriation, and adaptation, this research investigates how Shannon's work on information theory and its associated concepts, such as entropy and noise, were taken up and transformed as they traveled beyond the context of communication engineering into fields like computer science, linguistics, biology, and the social sciences. By attending to the ways in which these ideas were reinterpreted and repurposed in different disciplinary and institutional settings, we can gain insight into the broader intellectual and cultural significance of Shannon's work and its role in shaping the conceptual foundations of the information age.

To further explore the circulation of Shannon's ideas and their material instantiations, this research draws on methodological approaches from the field of STS, particularly those that emphasize the co-production of science, technology, and society. Building on the work of scholars like Sheila Jasanoff and Paul N. Edwards, this research examines how Shannon's theoretical insights were translated into concrete technologies and practices, such as communication systems, cryptographic techniques, and early computing machines, and how these material artifacts in turn shaped the social and institutional contexts in which they were developed and deployed. By tracing the reciprocal interactions between Shannon's ideas, their technological manifestations, and the broader social and political milieu of the Cold War era, we

can develop a more nuanced understanding of the historical forces that gave rise to the modern concept of data and its associated infrastructures.

Finally, this research engages with recent developments in the field of critical data studies, which seeks to interrogate the power dynamics, ideological assumptions, and ethical implications of data-driven technologies and practices. Drawing on the work of scholars like Lisa Gitelman, danah boyd, and Catherine D'Ignazio, this research attends to the ways in which data is never a neutral or objective representation of reality, but rather is always shaped by the values, interests, and assumptions of those who collect, process, and interpret it. By applying this critical lens to the history of data as a concept, we can examine how Shannon's work and its subsequent elaborations have both reflected and reinforced particular ways of knowing, categorizing, and acting upon the world, and how these epistemic and political commitments have been encoded into the technical infrastructures and institutional practices of the information age.

To operationalize these methodological and interpretive frameworks, this research will employ a range of specific research techniques and strategies. Close reading and textual analysis of Shannon's published works, correspondence, and other primary sources will be used to trace the development of his ideas and their expression in different rhetorical and disciplinary contexts. Archival research, including the examination of unpublished materials such as research notes, drafts, and working papers, will provide insight into the iterative process of conceptual development and the intellectual and institutional contexts in which Shannon's work took shape. Oral history interviews with Shannon's colleagues, collaborators, and students, as well as with scholars and practitioners who have been influenced by his work, will offer additional perspectives on the reception and impact of his ideas across different fields and historical periods.

In addition to these primary source-based methods, this research will also employ a range of secondary source analysis and synthesis techniques to situate Shannon's work within the broader intellectual and historical landscape of information science and adjacent fields. Literature review and bibliometric analysis will be used to map the citation networks and conceptual genealogies that link Shannon's work to subsequent developments in data science, computer science, and other domains. Comparative analysis of key concepts and frameworks across different disciplinary and national contexts will shed light on the ways in which Shannon's ideas were translated, adapted, and transformed as they circulated beyond their original milieu.

Ultimately, by combining these diverse methodological and interpretive approaches, this research aims to produce a rich and nuanced account of the intellectual history of data as a concept, one that attends to the complex interplay of social, institutional, and technological factors that have shaped its development and its ongoing significance in the contemporary world. Through a close examination of Claude Shannon's work and its afterlives, situated within the broader landscape of information science and its adjacent fields, this research seeks to contribute to ongoing debates and discussions about the nature, use, and implications of data in the 21st century, and to provide a historical foundation for the development of more critical and reflexive approaches to data-driven technologies and practices.

By employing a social constructionist framework, attending to the role of individual actors and institutions in shaping scientific fields, tracing the circulation of ideas across disciplinary boundaries, examining the co-production of science and technology, and engaging with critical perspectives on data and its associated power dynamics, this research aims to produce a multi-faceted and historically grounded account of the emergence and evolution of data as a key concept in modern information science and beyond. In doing so, it seeks to

contribute to the growing field of critical data studies and to provide new insights and tools for navigating the complex and rapidly changing landscape of data-driven knowledge production and decision-making in the 21st century.

V.   Historical Argument

Based on the research questions posed and the preliminary investigation of Claude Shannon's archive, this research puts forth the provisional argument that the universalization of data as a concept in the post-World War II era was fundamentally shaped by the intellectual and institutional context of early information theory and its intersections with the rise of digital computing and communication engineering. Shannon's groundbreaking work on the mathematical theory of communication, which introduced key concepts like information entropy, noise, and channel capacity, laid the conceptual foundations for a quantitative and statistical approach to information that would later be formalized as data science. However, the broader impact of Shannon's ideas and their subsequent development into a generalized notion of data cannot be fully understood without attending to the specific social, technological, and political contexts in which they emerged and circulated.

Central to this argument is the claim that Shannon's theoretical innovations, while deeply influential, did not arise in a vacuum. Rather, they were shaped by the intellectual milieu of Bell Labs and the larger field of communication engineering in the mid-20th century, which was grappling with the challenges of designing efficient and reliable systems for transmitting information across increasingly complex networks. Shannon's work was also informed by his engagement with the emerging field of computing and his collaborations with key figures like Alan Turing and John von Neumann, who were developing the theoretical and practical foundations of digital computation. These intersections suggest that the concept of data, as it

emerged from Shannon's work and its subsequent elaborations, was not a neutral or inevitable development, but rather a product of specific historical circumstances and intellectual trajectories.

To fully develop this argument, further research will be needed to trace the circulation and reception of Shannon's ideas beyond the immediate context of information theory and into adjacent fields like computer science, linguistics, biology, and the social sciences. This will involve examining how Shannon's concepts and methods were taken up, adapted, and transformed by researchers and practitioners in these domains, and how they intersected with broader intellectual and cultural currents of the post-war era, such as the rise of cybernetics, the growth of the military-industrial complex, and the emergence of new paradigms for scientific research and knowledge production. Additional archival research may also be necessary to uncover the institutional and political factors that shaped the development and dissemination of Shannon's work, such as the funding priorities of government agencies and private foundations, the strategic imperatives of the Cold War, and the changing landscape of intellectual property and research ethics.

Furthermore, to substantiate the claim that Shannon's work played a pivotal role in the universalization of data as a concept, it will be important to situate his contributions within the longer history of statistical and quantitative approaches to information and knowledge production. This will involve tracing the genealogy of key concepts and methods in fields like statistics, probability theory, and scientific measurement, and examining how they intersected with and informed the development of information theory and data science. By contextualizing Shannon's work within this broader historical narrative, it may be possible to identify the specific

intellectual and institutional factors that enabled his ideas to have such a profound and far-reaching impact on the conceptualization and operationalization of data in the modern era.

Ultimately, by pursuing these lines of inquiry and situating Shannon's work within the larger historical and intellectual landscape of the post-war period, this research aims to develop a more nuanced and historically grounded understanding of the emergence and evolution of data as a key concept in modern information science and beyond. While the provisional argument presented here provides a starting point for this investigation, further research and analysis will be necessary to fully test and refine these claims, and to explore their implications for ongoing debates and discussions about the nature, use, and governance of data in the 21st century.

Bibliography

Battle-Baptiste, W., & Rusert, B. (Eds.). (2018). W. E. B. Du Bois's data portraits: Visualizing

    Black America. Chronicle Books.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the New Jim Code. Polity

    Press.

Borgman, C. L. (2015). Big data, little data, no data: Scholarship in the networked world. MIT

    Press.

Bowker, G. C. (2005). Memory practices in the sciences. MIT Press.

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural,

    technological, and scholarly phenomenon. Information, Communication & Society,

    15(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

Buckland, M. K. (1997). What is a "document"? Journal of the American Society for Information

    Science, 48(9), 804-809.

    https://doi.org/10.1002/(SICI)1097-4571(199709)48:9<804::AID-ASI5>3.0.CO;2-V

Cheney-Lippold, J. (2017). We are data: Algorithms and the making of our digital selves. New

    York University Press.

Costanza-Chock, S. (2020). Design justice: Community-led practices to build the worlds we

    need. MIT Press.

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT Press.

Drucker, J. (2011). Humanities approaches to graphical display. Digital Humanities Quarterly,

    5(1), 1-21. http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Drucker, J. (2014). Graphesis: Visual forms of knowledge production. Harvard University Press.

Edwards, P. N. (2010). A vast machine: Computer models, climate data, and the politics of

    global warming. MIT Press.

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the

    poor. St. Martin's Press.

Ferraris, M. (2013). Documentality: Why it is necessary to leave traces (R. Davies, Trans.).

    Fordham University Press.

Foster, F. S. (2005). African American literary studies in the digital age: An introduction. African

    American Review, 39(3), 727-732. http://www.jstor.org/stable/40033686

Garcia Martinez, A. (2019, February 13). No, data is not the new oil. Wired.

    https://www.wired.com/story/no-data-is-not-the-new-oil/

Gitelman, L. (Ed.). (2013). "Raw data" is an oxymoron. MIT Press.

Hacking, I. (1999). The social construction of what? Harvard University Press.

Hunsinger, J. (2019). Big capta? In A. Romele & M. Terrone (Eds.), Digital Hermeneutics,

    Digital Methods, Digital Scholarly Editing (pp. 139-152). Springer International

    Publishing. https://doi.org/10.1007/978-3-030-37540-9_9

Isaacson, W. (2014). The innovators: How a group of hackers, geniuses, and geeks created the

    digital revolution. Simon & Schuster.

Jasanoff, S. (Ed.). (2004). States of knowledge: The co-production of science and social order.

    Routledge.

Johnson, J. M. (2018). Markup bodies: Black [life] studies and slavery [death] studies at the

    digital crossroads. Social Text, 36(4), 57-79. https://doi.org/10.1215/01642472-7145658

Klein, L., & Posner, M. (2020). Data. In B. Burgett & G. Hendler (Eds.), Keywords for

    American cultural studies (3rd ed.). NYU Press. https://muse.jhu.edu/book/75326/

Latour, B., & Woolgar, S. (1979). Laboratory life: The construction of scientific facts. Sage
    Publications.

Leonelli, S. (2016). Data-centric biology: A philosophical study. University of Chicago Press.

Lewenstein, B. V. (1995). From fax to facts: Communication in the cold fusion saga. Social
    Studies of Science, 25(3), 403-436. https://doi.org/10.1177/030631295025003001

Loukissas, Y. A. (2019). All data are local: Thinking critically in a data-driven society. MIT
    Press.

Manuscript Division, Library of Congress, Washington, D.C. Library of Congress. (n.d.). *Claude
    Elwood Shannon papers, 1932-1995 (Library of Congress Finding Aid)*.
    https://findingaids.loc.gov/db/search/xq/searchMfer02.xq?_id=loc.mss.eadmss.ms003071
    &_faSection=overview&_faSubsection=did&_dmdid=

McCarty, W. (2005). Humanities computing. Palgrave Macmillan.

McGlotten, S. (2016). Black data. In E. P. Johnson (Ed.), No tea, no shade: New writings in
    Black queer studies (pp. 262-286). Duke University Press.

Milner, Y. (2018). Abolish big data [Conference presentation]. 4S: Society for Social Studies of
    Science Annual Meeting, Sydney, Australia.

Monk, R. (1991). Ludwig Wittgenstein: The duty of genius. Free Press.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York
    University Press.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens
    democracy. Crown.

Patterson, S. (2015, October 2). Making data speak: The Colored Conventions archive unveils its
    new search tool. Colored Conventions Project.

https://coloredconventions.org/blog/making-data-speak-the-colored-conventions-archive-unveils-its-new-search-tool/

Posner, M., & Klein, L. (2017). Editor's introduction: Data as media. Feminist Media Histories, 3(3), 1-8. https://doi.org/10.1525/fmh.2017.3.3.1

Rogers, S. (2010, July 8). Data is the singular of datum (or why data is or are)? The Guardian. https://www.theguardian.com/news/datablog/2010/jul/08/data-plural-singular

Rosenberg, D. (2013). Data before the fact. In L. Gitelman (Ed.), "Raw data" is an oxymoron (pp. 15-40). MIT Press.

Secord, J. A. (2004). Knowledge in transit. Isis, 95(4), 654-672. https://doi.org/10.1086/430657

Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E. (1951). Prediction and entropy of printed English. Bell System Technical Journal, 30(1), 50-64. https://doi.org/10.1002/j.1538-7305.1951.tb01366.x

Wiggins, C., & Jones, M. L. (2023). How data happened: A history from the age of reason to the age of algorithms. W. W. Norton & Company.