DANIEL ROSENBERG*

Data as Word

ABSTRACT

The history of what we today call "data" extends to the ancient world, yet our contemporary terminology of "data" is modern. This article examines the history and significance of the term "data." It argues that a historiography of data that is self-conscious about the historicity of its own categories can illuminate the specific materiality of data, distinct from the things in the world it claims to represent. This essay is part of a special issue entitled *Histories of Data and the Database* edited by Soraya de Chadarevian and Theodore M. Porter.

KEY WORDS: data, eighteenth-century, words, language, materiality

"Data are things," writes historian of science Paul Edwards, and understanding data requires understanding the material and social systems of their production and circulation, their alignments and misalignments, and above all the different kinds of *friction* produced by the interaction of these materials and systems.[1]

Understanding that things called "data" are embodied and circulate in real, not just metaphorical, ways, opens the history of data to the full range of social and political questions posed by science and technology studies. In Edwards's book on climate modeling, the materiality of data is powerfully communicated in a photograph of the formal entry of the National Weather Records Center in Asheville, North Carolina, in the 1960s, clogged with

*Professor of History, University of Oregon, Robert D. Clark Honors College, 1293 University of Oregon, Eugene, OR 97403-1293, dbr@uoregon.edu

The following abbreviation is used: *OED*, *Oxford English Dictionary*.

1. Paul N. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (Cambridge, MA: MIT Press, 2010), 84; Paul N. Edwards, Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman, "Science Friction: Data, Metadata, and Collaboration," *Social Studies of Science* 41, no. 5 (Oct 2011): 669.

filing cabinets full of data cards, evoking not only data *friction* but data *mass* and *inertia* as well.[2]

While dictionary definitions of "data" suggest something immaterial—the *Oxford English Dictionary* defines "data" as "facts, esp. numerical facts, collected together for reference or information"—the closer we examine the way the term "data" is used, the more material it appears.[3] Consider, for example, the list of phrases that fall under the category "data" in the *OED*:

> data bank, data handling, data transfer, data transmission, data capture, data entry, data file, data link, data logger, data processing, data protection, data retrieval, data sheet, data structure, data terminal, data type, data analysis, data security, data acquisition, data communication, data dump, data smog, data stream, data warehouse, data suit, data glove[4]

In contrast to the *OED*'s covering definition of "data" as "facts," these phrases could hardly be better selected to buttress the argument that data are things. Collectively, the phrase list evokes the turbulent materiality of the contemporary information environment. In these compound phrases, the *OED* records data being logged, entered, handled, filed, banked, warehoused, and protected, when not dumped. Our visceral corporeal relationship to this data environment is equally well reflected in the "gloves" and "suits" enumerated by the dictionary, soon, I imagine, to be replaced by "implants." In recent years, a whole new wave of scholarship, including articles in the present collection, has filled out this picture.[5]

The problem of what we today call "data" is deeply material. The farther into the past one looks, the clearer this is, whether the material in question is

---

2. Edwards, *Vast Machine* (ref. 1), 102.

3. "Data, n." *OED Online*; http://www.oed.com.libproxy.uoregon.edu/view/Entry/296948 (accessed 7 Nov 2017).

4. Ibid.

5. See especially, Elena Aronova, Christine von Oertzen, and David Sepkoski, eds., "Data Histories," special issue, *Osiris* 32, no. 1 (2017); Lisa Gitelman, ed., *Raw Data (Is an Oxymoron)* (Cambridge, MA: MIT Press, 2013). On the historical layering of data materialities: Tung-Hui Hu, *A Prehistory of the Cloud* (Cambridge, MA: MIT Press, 2016); Shannon Mattern, *Code and Clay, Data and Dirt: Five Thousand Years of Urban Media* (Minneapolis: University of Minnesota Press, 2017). On the materialities of data on the Internet: Allison Carruth, "The Digital Cloud and the Micropolitics of Energy," *Public Culture* 26, no. 2 (2014): 339–64; Lisa Parks and Nicole Starosielski, eds., *Signal Traffic: Critical Studies in Media Infrastructure* (Urbana: University of Illinois Press, 2015). Some of this scholarship draws connections with an older historiography of record making exemplified by Jack Goody, *The Logic of Writing and the Organization of Society* (New York: Cambridge University Press, 1986).

paper, clay, or stone. Yet the notion there might be some things in the world that we ought to call "data," which undergirds much of the contemporary historiography on the subject, is historically new, and this is a fact that bears some examination. The presumption that a set of parish records, for example, should in the first instance be thought of as data, dates only to the eighteenth century, and even then, it was engaged with some tentativeness. Or, maybe better put, it was engaged with *purposefulness*.[6] It was not until much more recently, in the late twentieth century, that one came reflexively to refer to such information as "data."[7] To understand how this happened, it may be useful to know something not only about the history of data as a *thing* but also as a *word*. And, it may also be helpful to use some data on words to get at the problem.

In Latin, the term *data* is very common. It is nothing more than the plural of the neuter past participle of the verb *dare*, "to give." Thus, in the Latin Vulgate, Exodus 8:15 reads as follows: *Videns autem Pharao quod **data** esset requies ingravavit cor suum et non audivit eos sicut praeceperat Dominus*. "And Pharaoh seeing that rest was **given**, hardened his own heart, and did not hear them, as the Lord had commanded."[8] In all, the Vulgate uses the Latin word *data* more than fifty times, covering a striking variety of *things given* ranging from gold, silver, brides, fields, and towns to rights, power, glory, honor, and counsel. Yet, as early as the fourth century BCE, the Greek mathematician Euclid employed a specialized, technical usage of the term δεδομένα (*dedomena*) in Greek, or *data* in Latin. In Euclid, *data* are *givens* in a geometrical problem—typically points, lines, angles, polygons, and segments of circles—in contrast to *quaesita* or unknowns.[9]

6. On the purposeful usage of "data" at the start of the nineteenth century, see Christine von Oertzen's discussion of August Ludwig von Schlözer and Wilhelm Butte in "Datafication and Spatial Visualization in Nineteenth-Century Census Statistics," in this Issue.

7. It is only *very* recently that the implications of this shift in language have become fully apparent. Indeed, the shift is recent enough that it is neither registered in Raymond Williams's classic *Keywords: A Vocabulary of Culture and Society* (New York: Oxford University Press, 1976) nor even in the follow-on volume, Tony Bennett, Lawrence Grossberg, and Meaghan Morris, eds. *New Keywords: A Revised Vocabulary of Culture and Society* (Malden, MA: Blackwell, 2005).

8. Translation from the 1609 Douay-Rheims Bible.

9. The aspects of givenness in Euclid include position, magnitude, and form. Establishing what precisely Euclid means by *givens* is complex, particularly because Euclid has the habit of referring to *both* his premises and his conclusions as such. Taisback argues that this circularity is characteristic of both Euclid and of Greek ontology more generally. Christian Marinus Taisbak, "Elements of Euclid's *Data*," *Apeiron: A Journal for Ancient Philosophy and Science* 24, no. 4 (Dec 1991): 135–71. The seventeenth-century English translation of Euclid's *Data* includes a seven-page

In early modern Latin, the term *data* was still employed in the sense articulated by Euclid, often with direct reference to him. During this period, the term *data* mattered especially in the disciplines of mathematics and theology.[10] In both, *data* referred to things taken for granted and thus not inquired after. In early eighteenth-century texts, we can track minutely the transition of the term *data* from Latin to English, as for example, in William Pickering's 1710 *The marrow of mathematics*, which balances on the edge of Latin and English:

> . . . the taking of Heights and Distances at Sea, at Land, and in the Heavens, is the most principal, and the most practical part of the Mathematicks: And these are all perform'd by Trigonometry, or by the making and measuring, the framing and resolving of a Triangle, either Plane or Spherical: Supposing these Heights and Distances to be the several Parts, that is, either the sides or Angles of a Triangle. And to find by some of these Parts known, other Parts unknown or *per Data Quaesita*.[11]

In Pickering, *Data* are values entered into trigonometry problems—the *givens* in these problems—from which unknown values may be calculated. In this passage, Pickering indicates that in mathematics really existent things such as "heights and distances" are transformed into *data* that *do not* depend on the material world.

As is suggested by the definition of "data" given in the 1778 edition of the *Encyclopaedia Britannica*, as late as the end of the eighteenth century, Euclid's usage was still crucial. The *Britannica* puts it this way:

> From the primary use of the word *data* in mathematics, it has been transplanted into other arts; as philosophy, medicine, &c. where it expresses any quantity, which, for the sake of the present calculation, is taken for granted to be such, without requiring an immediate proof for its certainty. . . .[12]

As the *Britannica*'s definition shows, the concept, *data*, was from the start a powerful tool for argumentation, and its transplantation into fields other than mathematics was consequential. Euclid's approach facilitated consideration of hypotheticals. It cleared the ground for the formulation of theoretical

preface by the Greek philosopher Marinus (born ca. CE 440) explicating the term. *Euclid's Elements of Geometry* (London, 1661).

10. Daniel Rosenberg, "Data Before the Fact," in Gitelman, *Raw Data* (ref. 5), 15–40.

11. William Pickering, *The marrow of mathematics* (London, 1710), 7.

12. *Encyclopædia Britannica; or, a Dictionary of Arts, Sciences, &c.*, 2nd ed. (Edinburgh, 1778–83), vol. 4, 2382.

problems for which calculation was illuminating but the question of referential truth was, for one reason or another, moot.

A striking example of the power of the transplanted concept "data" occurs in early modern theology, and it is crucial to note that during the seventeenth and eighteenth centuries, it was in this field, above all, that Euclid's usage was generalized in English and other vernaculars. Thus, for example, John Shute Barrington's argument that theology is surer than science in *An essay on the several dispensations of God to mankind* (1728): "How much more then may we be assured of this in Revelation, where we have all the Data, before us?"[13] The notion that the facts of Revelation might be called "Data" feels strange today, but for a clergyman steeped in Latin learning, nothing was more natural. For Barrington, "Data" does not name one kind of thing or another. It identifies *whatever* is understood as given (in this case, the truths of Revelation).

As I have shown elsewhere, in theological disputation, "data" did not necessarily refer to *truths* in the way that Barrington uses the term here.[14] The term could just as well refer to claims accepted *for the sake of argument*. What made such claims "data" was their rhetorical status, not some intrinsic formal or material quality: in another argument, the very same statements might not be "data" at all. Theologians, like mathematicians, understood "data" to be whatever one treats as *given* in an analytic situation, whether geometric objects, numerical values, theological propositions, or something else again.

True, the term "data" was employed in several other fields in the seventeenth and early eighteenth centuries. William Petty, for example, used "data" in a discussion of ship fitting, though apparently not in his discussions of political arithmetic;[15] the *Proceedings of the Royal Society* employed "data" in discussions of astronomical calculations.[16] Yet, well into the eighteenth century, usage of the term in mathematics and theology predominated. In the first decade of the eighteenth century, approximately seventy percent of books that employ the term "data" pertain to the fields of mathematics or theology. In the last decade, the portion falls to about twenty percent.[17]

13. John Shute Barrington, *An essay on the several dispensations of God to mankind, in the order, in which they lie in the bible* (London, 1728), 2.

14. Daniel Rosenberg, "Data Before the Fact" (ref. 9), 20.

15. William Petty employed the term "data" in his *Treatise of Naval Philosophy* (London, 1691), 128, in order to stress the technical, mathematical aspects of his discussion of naval engineering: "Out of what *Data* arises the knowledge of the strength of the *Knees, Bolts,* and *Nails.*"

16. Royal Society of Great Britain, *Philosophical Transactions*, vol. 33 (London, 1726), 231.

17. Daniel Rosenberg, "Data Before the Fact" (ref. 9), 32–33.

The migration of the term "data" from Latin into vernacular languages in general took place mostly during the eighteenth and early nineteenth centuries.[18] In English, the importance of the term grew steadily during this period. At the start of the eighteenth century, the English word "data" appears in about one-third of one percent of English language books. At the end of the eighteenth century, it appears in about three percent, a ten-fold increase.[19] From that time, the frequency of use of "data" in English rises without interruption and by orders of magnitude.[20]

During the same period, the word "data" was naturalized into English. At the start of the century, "data" was italicized in more than eighty percent of its uses in English language books, indicating that it was treated as a foreign or a distinctive word. At the end of the century, "data" was italicized in less than twenty percent of uses.[21] That "data" became an English word during this period is confirmed also by the untethering of the term from the native grammar of Latin. In Latin, "data" is plural; in English, from the eighteenth century forward it is so only sometimes.[22] Today, one frequently encounters "data" used as a collective singular, in much the same manner as "media."[23] The emergence of the collective singular "data" is significant both from the point of view of grammar and from that of semantics. From the eighteenth century forward, the English language provides tools for discussing "data" as *something*.

Furthermore, during the seventeenth and eighteenth centuries, the domains in which the term "data" was employed multiplied and shifted. At the start of the century, usages in mathematics and theology were most important; at the end of the century, the term "data" was employed increasingly in fields with which we commonly associate it today, including population studies and finance.[24]

18. Ibid., 30–31.

19. Ibid., 28–29.

20. In 1800, the term "data" accounts for 0.0004 percent of the words in the Google Books corpus according to the Google Books Ngram Viewer. In 1900, it accounts for 0.0016 percent of the corpus. In 2000, it accounts for 0.0324 percent. http://books.google.com/ngrams (accessed 7 Nov 2017).

21. Rosenberg, "Data Before the Fact" (ref. 9), 32–33.

22. In addition to Vauban, examples include Edward Laurence, *The duty of a steward to his Lord* (London, 1727), and George Adams, *The description and use of the universal trigonometrical octant* (London, 1753).

23. I discuss the conjugation of "data" in Rosenberg, "Data Before the Fact" (ref. 9), 18–19. In my view, there is neither grammatical nor historical merit to the argument that "data" must always be plural in English.

24. Ibid., 30–31.

Across disciplines, the subject of data was gradually spoken about in new ways. An analysis of noun/verb collocations shows that at the beginning of the nineteenth century, data were still mostly "afforded" or "furnished." At the beginning of the twentieth century, data were "collected" and "published." At the start of the twenty first century, they were "entered," "transmitted," and "received."[25]

At the end of the eighteenth century, the term "data" was not ubiquitous in the way that it is today. "Data" remained mostly a term of art, and the *givenness* that it implied depended on the arena in which it was employed. In medicine, "data" typically referred to the direct observations of the physician; in psychology, to sense impression; in philosophy, to first principles. Regardless of domain, the term's use pointed explicitly (and often polemically) to what the writer proposed as a standard of givenness.

To understand how "data" worked in early modern English, consider, for example, a fascinating, little-remembered text on magnetism published in 1761 by the fascinating, little-remembered experimentalist, Temple Henry Croker, which makes the following surprising statement: "Till Experimental Philosophy was introduced, All Science was founded upon Data."[26] Without some historical context, it is hard to guess what Croker could possibly have meant by this. Today, one would expect a statement more like: *Not until* Experimental Philosophy was introduced, *was* Science founded upon Data. Yet, according to Croker, experimental philosophy represents scientific advance because it *rejects* reasoning upon "Data".

Croker was no great scientist. Despite some successes including the design of a horizontal windmill that is said to have functioned, his legacy is meager.[27]

25. Collocation analysis performed with Mark Davies, *Google Books Corpus* (based on Google Books Ngrams), 2011–. http://googlebooks.byu.edu/ . Another tantalizing pattern in the Google Books corpus: when they talk about data, earlier texts are more likely to refer to a *lack* of data than later texts. Over time, the frequency of phrases such as "insufficient data," "want of data," or "no data" declines in relation to occurrences of the term "data." Given the history of data production during these centuries, this is not entirely surprising, but it is suggestive. An interesting example consistent with the pattern: Thomas Malthus uses "data" only once in *An essay on the principle of population* (London, 1798), 71–72: " . . . to prove distinctly the force and extent of its operation, would require, perhaps, more data than we are in possession of."

26. Temple Henry Croker, *Experimental Magnetism, or, the Truth of Mr. Mason's Discoveries* (London, 1761), 7.

27. Temple Henry Croker, *Advantages proposed by Croker's invention on an horizontal windmill for grinding sugar-canes* (Basseterre, St. Christopher, 1784). Croker was also editor of a three-volume *Dictionary of Arts and Sciences* (1766), translator of a small number of literary works, and author of several sermons and works of criticism.

A manuscript note on Croker's letter to Benjamin Franklin promoting his project rejects it out of hand as "published . . . long ago" and "founded on a Mistake."[28] Yet Croker's statement about data was less strange than it appears in retrospect. Croker did not maintain that experimental philosophy should reject data in the sense that we employ the term today. He used "data" to mean *unfounded assumptions*, and his critique of data in this sense closely resembles contemporary critiques of the *spirit of system* as, for example, in the *Encyclopédie* of Denis Diderot and Jean le Rond d'Alembert. Croker writes:

> The Discovery of Truth must ever be agreeable, and in many Sciences, especially that of the Mathematicks, useful to the Community. The Splendor of Truth may offend the weak Eyes of Individuals, but must comfort and cheer those of the Publick; and though it should happen that the Authority of great Names, or the long and general Acquiescence in the Laws of a Favourite System, should for a while with-hold their Consent, yet Truth, founded on Facts (and Experiments can go by no other Name) must at last emerge triumphant over the Waves of Opposition, and the vain Attempts of Fancy to support an Hypothesis, dictated by itself, and settled upon no stronger a Foundation. Till Experimental Philosophy was introduced, All Science was founded upon *Data*. Give the system-maker's their first *Principia*, they then proceed boldly, from them to account for every Appearance; but deny their first Position, and their Science is at an End.[29]

Further scientific advance, Croker writes, "must result, not from Fancy but from Facts, not from artfully devised Systems, but from real Experiments."[30]

Odd as it seems, for Croker, as for many other writers in his day, "data" still meant *things given*, whether facts or principles. In a disputation with John Wesley from 1759, for example, John Free explicitly defines "data" to mean *axioms* or *maxims*:

---

28. Letter from Temple Henry Croker to Benjamin Franklin, Philadelphia, 11 June 1787, http://www.franklinpapers.org/ (accessed 7 Nov 2017). A hand-written note on letter dated 15 June 1787 reads: "The substance of this paper has been published in a pamphlet long ago. All the reasoning on the subject is founded on a Mistake of the Laws of Composition and Resolution of forces."

29. Croker, *Experimental Magnetism* (ref. 26), v–vi.

30. Ibid., 71–72. Twenty years later, for example, we find Croker's formula inverted almost exactly by John Berkenhout: "a modern politician therefore, instead of a Puffendorf, a Locke, a Sydney, a Montesquieu, must shut his books, forget his reading, and, on such a data as the world before him will afford, dare to reason for himself." *Lucubrations on ways and means: Humbly addressed to the Right Hon. Lord North* (London, 1780), 6.

> The English reader is to understand, that in every Dispute, there are some Things *granted* on both Sides, while others remain *to be determined*. The Things *granted* are called in the Greek $\Delta\varepsilon\delta o\mu\acute{\varepsilon}\nu\alpha$, in the *Latin* DATA, by which Name, through the Disuse of our own Words, they are sometimes called in *English*; or else *Axioms* or *Maxims*; as Things of prime Evidence, and *worthy of all Men to be received*.[31]

We find a similar usage in an amusing passage in Tobias Smollett's satirical novel *Peregrine Pickle*, in which Peregrine's tutor, Mr. Jolter, attempts to construct a mathematical proof of the importance of romantic virtue. Jolter,

> . . . hemming thrice, observed, that no mathematical inquiries could be carried on, except upon certain *data*, or concessions of truths, that were self-evident; and therefore he must crave his assent to a few axioms, which he was sure Mr. Pickle would see no reason to dispute.[32]

These included the statements that "youth and discretion . . . remain still equi-distant, and will never coincide" and that "passion acts upon the human mind, in a ratio compounded of the acuteness of sense, and constitutional heat."[33] To Free and Smollett, as to Croker, "data" meant things assumed, not those discovered by investigation or experiment.

In the mid-eighteenth century, however, the sense represented in these passages was waning. The meaning of the word "data" was not changing: it still meant *things given*, but in the context of eighteenth-century empiricism, assumptions about what constituted *givenness* were changing importantly in ways that laid the foundations for our contemporary usage.

By the second half of the eighteenth century, a modern vocabulary of "data" was emerging. If the language of Croker, Free, and Smollett strikes us as unfamiliar, the same is not true for a passage we find in a 1775 letter by Benjamin Franklin, which uses the term "data" with some irony, to describe a calculation of costs and benefits of the war between the British armed forces and the rebellious English colonists in North America. Franklin writes,

> Tell our dear good friend [Dr. Price], who sometimes has his doubts and despondencies about our firmness, that America is determined and unani-mous; a very few tories and placemen excepted, who will probably soon

31. John Wesley, *Dr. Free's Edition of the Rev. Mr. John Wesley's Second Letter* (London, 1759), 3–4.

32. Tobias Smollett, *The Adventures of Peregrine Pickle*, vol. 1 (London, 1751), 186–87.

33. Ibid., 187. On the presence of "data" in eighteenth-century imaginative literature, see also Daniel Rosenberg, "Whence *Data*?" *Berlin Journal* 28 (Spring 2015), 22.

export themselves. Britain, at the expence of three millions, has killed 150 Yankies this campaign, which is £20,000 a head; and at Bunker's Hill she gained a mile of ground, half of which she lost again by our taking post on Ploughed Hill. During the same time 60,000 children have been born in America. From these data his mathematical head will easily calculate the time and expence necessary to kill us all, and conquer our whole territory.[34]

Here, "data" refers to quantitative facts gathered through observation, collection, and experiment, which are then subject to mathematical manipulation and scientific or social analysis. That Franklin might use the term so casually in a letter, and with irony, suggests that he took his usage to be transparent, if perhaps still novel. Croker's evident labor in employing the term suggests at least a hint of the contrary.

During the eighteenth century, then, a general shift took place by which the term "data" came to be associated with many topics and concerns familiar today. But the key shift of interest here has only secondarily to do with subject matter. The more important point concerns the emergent positivity implied by the term. Franklin's roving, ironic commentary draws upon a general fund of—as it happens, quantitative—information that demands a name. The fact that "data," a term that had functioned to delineate a kind of counter-positivity, fit the bill is in many ways ironic. And yet it did: the term "data" came to evoke a particular sort of representational entity upon which one could operate through systems of calculation, classification, and communication, while holding the question of referential truth in abeyance.

As the definition of "data" in the 1778 *Britannica* makes clear ("any quantity, which, for the sake of the present calculation, is taken for granted to be such, without requiring an immediate proof for its certainty"), the connection between the usage we find in Franklin and the older usage of "data" remained important throughout the eighteenth century, even as "data" came more to be understood as the kind of *thing/s* of which the historiography in this volume seeks to make an account.

The utility of talking about "data" in this way is not hard to perceive, and the rapidly growing use of the term in the scientific, bureaucratic, and popular

34. Benjamin Franklin, from Letter to Joseph Priestley, 3 Oct 1775, in Benjamin Vaughan, ed., *Political, Miscellaneous, and Philosophical Pieces . . . by Benj. Franklin* (London, 1779), 365–66. Richard Price, referenced here by Franklin, too, used "data" in interesting ways. In his *Four dissertations* (London, 1767), 25–26, "data" to refers to the visible work of God. In his *Observations on the expectations of lives* (London, 1769), 38–39, it refers to population counts in parish registers.

literature of the late eighteenth century confirms that it was not obscure to contemporaries. At the same time, the utility of the concept should obscure neither its contingency nor the specifically rhetorical problem that accounts for its emergence in the first place. A challenge for us is to retain the critical perspective on "data" that was common during much of the eighteenth century, as well as a sense of the historicity of our own enterprise.

ACKNOWLEDGEMENTS