Why Digital Humanists Should Emphasize Situated Data over Capta

Matthew Lavin <lavinm_at_denison_dot_edu>, Data Analytics Program, Denison University

Abstract

This essay looks back on Johanna Drucker's "Humanities Approaches to Graphical Display" (2011) ten years after its initial publication in *Digital Humanities Quarterly*, in particular Drucker's call to "reconceive all data as capta." Drucker makes several crucial points about humanistic inquiry, but this essay argues against her embrace of capta as a replacement term for data in two ways: (1) furnishing a revised and expanded etymology for the terms data and (2) exploring the benefits of embracing concepts such as situated data rather than capta.

As we pass the ten-year anniversary of Johanna Drucker's "Humanities Approaches to Graphical Display" (2011), I believe it is an opportune time to revisit Drucker's call to "reconceive all data as capta" [Drucker 2011].[1] Drucker's argument for such a reimagining is based on "the etymological roots" of data and capta — data comes from the Latin "given," and capta from the Latin "taken" — and raises a larger point about the differences between scientific realism and constructivist critiques of realism.[2] Realist approaches to visualization, Drucker argues, assume "transparency and equivalence, as if the phenomenal world were self-evident and the apprehension of it a mere mechanical task" [Drucker 2011]. Drucker goes further in stating, "Nothing in intellectual life is self-evident or self-identical, nothing in cultural life is mere fact, and nothing in the phenomenal world gives rise to a record or representation except through constructed expressions" [Drucker 2011]. Drucker lays out a lexical argument between data and capta that more or less parallels the realist and constructivist modalities of thought as she describes them (i.e., data suggests realism and capta suggests constructivism).[3]

1

In this piece, I want to argue in favor of embracing some of Drucker's points about humanistic inquiry while simultaneously arguing against capta as a term to be used in place of data. As a humanist, I do see value in emphasizing that data are taken and not given, but I believe there is a richer etymological narrative, and a richer history of the word *data* in English, to be described. There is also the more complicated question of whether all forms of empiricism require proceeding "as if the phenomenal world were self-evident," and the degree to which social constructions mediate our experiences and understanding of the world. I will gesture at these larger questions in this essay, but my primary concern is with how scholars in digital humanities should approach conversations about data.

2

The bulk of my essay is a revised and expanded etymology for the terms *data* and *capta*, which considers their Latin roots, as well as the close ties of the word *data* to the publishing history of *Euclidis Data* (1625) and subsequent translations and editions of that work. Euclid's own use of the term "given" was a subject for discussion and debate, as was the degree to which his use of "given" resembled other Greeks' use of the term. A fully fleshed out history of the term *data* in English includes its use in two related but distinct senses: geometrical and empirical, the latter of these emerging from the prior. To appreciate the difference between these senses, we must understand the degree to which early modern thinkers synthesized geometrical and numerical thought into what is now regarded as the singular discipline of mathematics. Further, these thinkers adapted and extended mathematical thinking to translate observations of or measurements from the natural world into generalizations and explanatory systems.

3

In the closing section of this essay, I will turn to the benefits of embracing concepts such as situated data over capta. Such an approach allows humanists to contest oversimplifications of poorly executed data-driven inquiry and

4

simultaneously to create more opportunities for conversation with other disciplines. The tone of these conversations could be positively affected, as well, with digital humanities speaking and listening in equal proportion. Lastly, I will discuss the new possibilities that this rhetorical shift could create for how we teach data analysis, which has the potential to advance digital humanities as a discipline.

Drucker's Conceptual and Terminological Intervention

My response to Drucker begins with a certain degree of uncertainty about how "Humanities Approaches to Graphical Display" has been received in digital humanities. Generally, I take it to have made a strong, positive impression. Drucker was not the first scholar to point out that *capta* might be a more appropriate term than *data*, but her work is arguably the most influential to make this point, especially in digital humanities.[4] This essay, or the corresponding material in Drucker's book *Graphesis* (2014), is widely referenced in digital humanities, and I have seen it listed on many DH course syllabi.[5] There appear to have been numerous conference presentations, blog posts, and discussions of capta on social media. In various contexts, public displays of one's agreement with Drucker might take the form of statements like, "There is no such thing as 'data.' There are only 'capta.'"[6] Broadly speaking, Drucker has set the terms of the debate on the subject of data and capta, and this influence has been both descriptive and normative. Simultaneously, I have not seen signs of a large-scale movement to purge documents of the word *data*, nor have the word *capta* or its cognates become especially prevalent in digital humanities scholarship.[7]

One might argue that Drucker makes a conceptual intervention and not a terminological one. I would concede that the essay is most concerned with reconceptualizing data as capta, and the importance of such a shift. With regard to data visualization, Drucker argues that "the rendering of statistical information into graphical form gives it a simplicity and legibility that hides every aspect of the original interpretative framework on which the statistical data were constructed" [Drucker 2011]. Data do not "pre-exist their parameterization" because they are "constructed as an interpretation of the phenomenal world, not inherent in it" [Drucker 2011]. Drucker does not explicitly argue that humanities scholars should never say or write the word *data*, but she does claim that "all data is capta" nd that because of the etymological roots of the terms, *capta* "is 'taken' actively while *data* is assumed to be a 'given' able to be recorded and observed" [Drucker 2011]. As a result, I reject the idea that one can so easily separate Drucker's conceptual intervention from the terminological one supporting it.

Another way of looking at Drucker's influence would be to say that there is widespread agreement with her views on data and capta, but practical considerations outweigh other concerns. Such practical considerations include (1) maximizing clarity for particular audiences, (2) limits on space that would make it difficult to explain the choice to use *capta*, and (3) the question of how to invoke concepts like big data, data visualization, databases, datasets, metadata, and open data. One might attempt to stake out a rhetorical middle position by mentioning that data are really capta and to proceed with the word *data* thereafter, as Rob Kitchin does in *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (2014). Kitchin adds, "since the term data has been so thoroughly ingrained to mean capta, rather than confuse the matter further it makes sense to continue to use the term data where capta would be more appropriate" [Kitchin 2014]. This seems like it would be an appealing option for many people, and I suspect it is widespread, although I have not investigated this question in any serious way.

Regardless, I am aware of no published work that disputes Drucker's core points. That is:

- the word *data* connotes something given and
- reconceptualizing data as capta is integral to rejecting realist assumptions of transparency and equivalence between data and the phenomena they purport to represent.

A Revised and Expanded Etymology for Data and Capta

In direct response to Drucker, I hope to demonstrate that the word *data* does not assume a given that can be transparently and unproblematically recorded and observed. If this first point can be satisfactorily established, it should follow that one can say and write the word *data*, embrace "the situated, partial, and constitutive character of knowledge

production," and recognize that "knowledge is constructed, *taken*, not simply given as a natural representation of pre-existing fact" [Drucker 2011]. This line of reasoning is crucial to how digital humanists understand and talk about the concept of data, as well as how we teach computational methods. However, making such an argument first requires revisiting the origin of *data* as a term of use in English and discussing how the meaning of the term *data* has changed over time.

To begin, the word *data* indeed comes from the Latin for "given," but the etymology of the English language word is less straightforward than Drucker suggests. A more detailed account of "the early history of the concept of 'data'" is Daniel Rosenberg's "Data Before the Fact" (2013), which discusses much of the early history of data's English usage. As Rosenberg describes, "In English, 'data' was first used in the seventeenth century. Yet it is not wrong to associate the emergence of the concept and that of modernity. The rise of the concept in the seventeenth and eighteenth centuries is tightly linked to the development of modern concepts of knowledge and argumentation" [Rosenberg 2013, 15]. Rosenberg emphasizes the "specifically rhetorical" status of the word *data* as something both "pre-analytical" and "pre-factual," more so even than the related concepts of "facts" and "evidence" [Rosenberg 2013, 19]. As he further explains, "When a fact is proven false, it ceases to be a fact. False data is data nonetheless" [Rosenberg 2013, 18]. Rosenberg's telling of the early history of data in English adequately distinguishes between the two earliest senses of the term *data*, the first in "the realm of mathematics, where it retained the technical sense that it has in Euclid, as quantities *given* in mathematical problems, as opposed to the *quaesita*, or quantities *sought*" and the second "in the realm of theology, where it referred to scriptural truths — whether principles or facts — that were given by God and therefore not susceptible to questioning" [Rosenberg 2013, 19]. However, I want to argue that these two senses are more closely related than one might first assume, and there is a third early sense of the word that Rosenberg does not fully distinguish.

First, the *Oxford English Dictionary* provides an entry for data as "something given or granted; something known or assumed as fact, and made the basis of reasoning; an assumption or premise from which inferences are drawn" [OED 2020c]. We might call this "data in the geometrical sense." This phrasing is intended to echo Rosenberg's definition, except for the distinction that our contemporary concept of mathematics depends on a synthesis of geometrical and numerical paradigms of thought that had not yet fully unfolded in Europe circa 1600.[8] The story of the data's entry into English must include some history of the translation of Euclid's work from Greek to Latin and English, a story Rosenberg alludes to but does not detail in full. The Greek word δεδομένα (dedoména) means "given." Δεδομένα (*Dedoména*) is the title of a work by Euclid, which remains extant in manuscript form in Greek and Arabic. It was translated into Latin by the French mathematician Claude Hardy and printed alongside its original Greek as a volume titled *Euclidis Data* in 1625. An English translation with the title *Euclid's Data* was included with an edition of *Euclid's Elements of Geometry* (John Leeke & George Serle) in 1661.[9] The *OED* traces this sense of the word *data* to Henry Hammond's *A copy of some papers past at Oxford, betwixt the author of the Practicall catechisme, and Mr. Ch.* in 1646 [OED 2020c], which refers to "a heape of data," but, as Rosenberg states, this heap "is not a pile of numbers but a list of theological propositions accepted as true for the sake of argument" [Rosenberg 2013, 20]. It may seem counter-intuitive to associate the Greek-to-Latin geometrical usage of *data* with what Jonathan Furner calls "data as gifts from God," but the theological or ecclesiastical notion of data is a direct antecedent of how Euclid used the term *data*/*dedoména* [Furner 2016, 292]. Furner elaborates that such uses go back to Thomas Tuke's *Nevv Essayes* of 1614, and that the Latin word *data* is found in other Latin phrases in "religious texts of early seventeenth-century England — for example, *gratia gratis data* ('grace freely given'), and *data desuper* ('given from above')" [Furner 2016, 292]. Like Rosenberg, Furner differentiates between *data* in the geometrical sense and *data* in the ecclesiastical sense, but they both owe their meaning to a fairly literal translation of the Latin word as "given."

Quite different from this first cluster is what we might call "data in the empirical sense." The *OED* refers to datum "chiefly in plural" (i.e., data), which describes "an item of (chiefly numerical) information obtained by scientific work, a number of which are typically collected together for reference, analysis, or calculation" with first use listed as 1630 [OED 2020c]. [10] This sense of the word seems to relate to the *OED*'s entry for data "as a count noun: an item of information; a datum; a set of data" (first use listed as 1645) and later, as "a mass noun," meaning "related items of (chiefly numerical) information considered collectively, typically obtained by scientific work and used for reference, analysis, or calculation"

with its first use listed as 1702 [OED 2020c].[11] There are likely examples of historic usage where data appears to be something in between the geometrical sense and the empirical sense, but the difference between the two in their sharpest forms is quite significant, and assuming the associations of the first sense carry over to the second sense would be a mistake. To better understand the dissimilarities between these senses, we must reexamine the first English translation of Euclid's *Data*.

How Euclid had used *dedoména*/*data* is the subject of a commentary by Marinus of Neapolis, translated and included in Leeke and Serle's edition, for "the Ancients have defined it in one manner, and later Writers after another" [Marinus 1661, 533]. For Marinus, the question was whether Euclid's notion of data could be defined as a combination of *ordinatum* (regulated or orderly) and *poriman* (available or provided) or as a combination of *cognitum* (infamous or known) and *porimon*. Marinus' two candidate definitions may appear to map directly to the English language senses of geometrical data and empirical data, but this is not the case. Neither of Marinus' notions would have been used as a term to describe aspects of the natural world, only for abstractions like the following:

1. The length of a line C when line C equals the length of line A minus the length of line B, and the lengths of lines A and B are both given.
2. The length of line Z connecting two points A and B, when the positions of points A and B are given.

Drawing on examples from Euclid and other ancient texts, Marinus concludes that Euclid is referring to data as a combination of *porimon*, "that which may be exhibited by Demonstration, or which is apparent without Demonstration" and *cognitum*, that which is "clear and comprehended of us" [Marinus 1661, 535, 534]. However, his recognition of a second candidate definition, describing data as a combination of *ordinatum* and *poriman* should be noted, as it could point to others forming a similar impression.

More recently, Christian Taisbak notes that Euclid's use of the word *data* means both the clear and comprehended premise, and the demonstrated conclusion:

> When I started to translate the *Data*, I found it very longwinded that a certain phrase kept popping up time and again, several times in every proposition: *if this item is given, that item is* also *given*. I decided to cancel all those *alsos* and restore them only where they were absolutely necessary. But then I discovered that I was leaving out an essential feature of the *Data*: the Givens hang together in chains, the purpose of any proposition being to produce more links to them. [Taisbak 2003, 14]

Taisbak is essentially describing deductive reasoning, "that if some items are given, some other items are also given, *into the bargain* so to speak" [Taisbak 2003, 13]. For Euclid, *dedoména* referred "not only to the input of a problem, but also to the output" [Taisbak 2003, 13]. In this context, one might also compare Euclid's use of *data* to *akolouthia* (ἀκολουθία), a term favored by Aristotle and others, which "indicates the necessary relationship between two propositions when one of them is the consequence of the other" [Seco et al. 2010 , 15]).[12] This perspective complicates Rosenberg's gloss of *data* as quantities given and *quaesita* as quantities sought, since quantities sought can in fact become given through demonstration or deduction (ὅπερ ἔδει δεῖξαι / *hóper édei deîxai*).[13] These nuances remain specific to the geometrical rather than the empirical concept of data that I have so far discussed, but they are important aspects of the history of the term *data*, as well as the history of how that term became associated with quantitative reasoning. As a result, these details already suggest the beginnings of a pedagogical intervention in digital humanities, where the "given-ness" of data is more complex that it might initially appear.

Data in the empirical sense appears to draw upon aspects of data as a geometrical given and extend these qualities to features of the natural world. For the likes of Apollonius, Euclid, and Marinus, the concept of data would never imply collecting information in this way. Observations of the sort said to be described in Eratosthenes' *Γεωγραφία* (*Geografíka*), would be called φαινόμενο (*fainómeno*), or "phenomenon," which translates to "that which appears or is seen."[14] Phenomena included anything that appeared to be true, such as illusions, mirages, and dreams. James Evans and J. Lennart Berggren explain that "the word 'phenomena' is a participle of the passive verb 'phanomai', which carries the meanings of 'to come to light, come to sight, be seen, appear.'" They add, "The last two are definitive for the

astronomical sense of the word, which is 'things that are seen/appear in the heavens'" [Evans and Berggren 2018, 5]. An entire genre of Greek philosophical thought was dedicated to the idea that, from careful observation of the heavens, one could understand the positions of the earth, the planets, and the stars, as well as explain the circular motions of all celestial bodies. Hence the notion, originating in Greek astronomy, of σωζειν τα φαινόμενα (sozein ta fainómena) or "save the phenomena" [Heath 1921, 7]. Works within this genre included Euclid's *The Phenomena*, Autolykos' *On the Moving Sphere*, Aristotle's *On the Heavens*, Gemino's *Introduction to the Phenomena*, and Theodesius' *Sphaerics*, which is thought to extend the work of Eudoxos. Contributions to this genre continued into the Middle Ages.

These writers did not appear to distinguish between their observations and the way they were structured or recorded. Much later, the genre of "Almanack of Ephemeris" or "Ephemerides" seems to suggest a potential term that makes this distinction. The earliest use of the term "almanac", meaning "an annual table, or (more usually) a book of tables," can be traced to *The Equatorie of the Planetis* (circa CE 1392), whereas the term *ephemeris* was used as early as the mid-16th century [OED 2020a]. According to Arthur L. Norberg, an almanac of ephemeris, by the nineteenth century, "contained information on the phenomena for that year: eclipses of the Sun, Moon, and Jupiter's satellites; the orientation of Saturn's rings and the apparent discs of Venus and Mars" [Campbell-Kelly 2003, 196]. For whatever reason, neither of these terms caught on as a more generalized word for the kind of structured information one might generate by making and recording observations. As with Taisbak's point about givens, for Euclid, functioning as the inputs and outputs of problems, the seeds of a pedagogical intervention are here planted. Phenomena were recorded and published with the expectation that future observations might supplement, refine, or replace previous ones, which suggests a fundamental awareness that representations of phenomena were partial and imperfect. Certainly, at almost any moment in history, some were expressing their belief in a unified, coherent reality just beyond our reach, as well as their confidence that their tabulations would allow them to decipher its governing dynamics. I will return to this subject in this essay's conclusion but, for now, it should suffice to say that this belief is best understood not as a position on the distinguishing properties of data, but rather a position on the possibility of obtaining objective knowledge through empirical means.

It is also crucial to note that the idea of a table as a textual and material device for structuring numerical information is at least 4,500 years old. According to Martin Campbell Kelly and his co-authors, "While the list has been hailed as a major breakthrough in cognitive history...the table as a pre-modern phenomenon of structured thought has been completely neglected" [Campbell-Kelly 2003, 13]. In ancient Sumer, Babylonia, and Assyria, clay tablets were often if not primarily a repository of numerical information, often used as a memorization or calculation aid. Etymologically, the word "table" comes from the Latin "tabula", which in turn, originates with the Greek term "τάβλι" ("tabla"), meaning a plank or board [Austin 1934, 202–5]. The term was used to refer to a particular board game that is thought to be the ancestor of backgammon, and was also a general term for a tablet, a slate, or any flat piece of wood. In English, it eventually a particular piece of furniture "on which food is served, and at or around which people sit at a meal" (circa CE 1300). The English word "table" meaning "a schematic arrangement of information" or "an orderly arrangement of particulars" can be traced as far back as Byrhtferth of Ramsey's *Enchiridion* (CE 986-1016), but the association between a tablet/slate and structured information recorded upon it appears to be much older [OED 2020e].

Today, the phrase "tabular data" suggests a strong association between tables and data, and the term "observation" is widely used to denote "all values measured on the same unit (like a person, or a day, or a race) across attributes" or, in the context of rectangular data, a row of values.[15] As Lisa Gitelman notes, the idea of converting that which one observes into data — a conceptual object that structures and represents that which has been perceived — seems like second nature to us. "When phenomena are variously reduced to data, they are divided and classified, processes that work to obscure — or *as if* to obscure — ambiguity, conflict, and contradiction" [Gitelman 2013, 9]. For Gitelman, this relationship suggests both a dependency on hierarchy, and a kind of epistemological cover that protects "users who perform logical operations on the data...from having to know how the data have been organized" [Gitelman 2013, 9]. When terms like "data-driven" are invoked, there is often an assumption of order, imposed or revealed, by those engaging in analysis.

Nevertheless, the idea of ordering and standardizing information for the purposes of facilitating numerical operations is much older than the English word *data*. What seems to have changed is that the term *data* adopted a secondary meaning in addition to the geometrical sense of term associated with Euclid. An entry in Ephraim Chambers'

Cyclopædia (1728) offers two definitions for the term *data*. First, it continued to refer to a geometrical given. Second, "From the primary Use of the Word *Data* in mathematicks, it has been transplanted into other Arts; as Philosophy, Medicine, &c. where it expresses any Quantity, which for the Sake of a present Calculation, is taken for granted to be such, without requiring an immediate Proof for its Certainty" [Chambers 1728, 64]. According to Chambers, data had come to be used to describe a quantity, in a wide range of contexts, that was meant to be treated as a given for the purposes of analysis. This definition reinforces the distinction between data in the strictly geometrical sense and data in other contexts, and it is compatible with religious uses of the term, since information from scripture would qualify as data of the geometrical sort, or knowledge beyond question.

This explanation does not describe a specific process by which phenomena become data. However, the clause "for the Sake of a present calculation" seems to limit why and how the data are permitted to be taken for granted, which undercuts the idea that data would be regarded as natural or pre-existing facts. In turn, it contradicts the idea that data in this sense could ever be viewed as pre-factual, as Rosenberg describes. Among the natural philosophers of the 17th and 18th centuries, the idea that Nature would reveal its hidden structure or laws to those who spoke its language was predominant.[16] In this light, it seems likely if not obvious that empirical observations would be considered facts given by God. However, if data were available or provided because Nature had provided them for us to analyze, then there would be no need for Chambers' qualifying phrase. The entry in Chambers' *Cyclopædia* seems more compatible with the idea that phenomena become data, or are granted provisional status as data, which suspends the need for "an immediate proof" and permits calculation to occur.[17]

Returning to Marinus' two candidate definitions for Euclid's notion of data — as *porimon* plus *cognitum* or *porimon* plus *ordinatum* — we can see a case for either or both.[18] Imposing Marinus' analysis on Chambers' definition is potentially problematic, but I would argue that our contemporary use of the term *data* is best explained by starting with the notion of data as *porimon* plus *ordinatum*. Recall that this sense — as a combination of something provided in a way that gives it order — was *not* the sense Euclid was not using, according to Marinus. As Taisbak notes, however, *ordinatum* for Marinus was the Greek word τεταγμένων (*tetagmenon*), literally "fixed," but also used to mean "organized according to some *taxis*, 'order'" [Taisbak 2003, 242]. Regarding the Latin *porimon*, the Greek term πορίσαθαι (*porizesthai*), Taisbak explains that Euclid uses the term to describe "several operations," including "put together," "draw," "describe," "make," or "produce" [Taisbak 2003, 242].[19] By this logic, defining *data* as both *porimon* and *ordinatum* would emphasize its being available, collected, or made, as well as being brought to order or revealed to have latent order by the act of demonstration.

Broadly speaking, then, a full-fledged etymology of the term *data* suggests a narrative much more complex than data as "given" and capta as "taken." The Latin word for "given" was used by Euclid in a manner very different from how data was used by the mid-18th century, and the true meaning of the Euclidian concept of dedoména/data has disputed definitions among classical thinkers, as Marinus summarizes. Concepts such as saving the phenomena, an almanac of ephemeris, and tabular information suggest an ancient and longstanding tradition of making precise observations, placing them a structured format, and using that structure to facilitate calculation. Taking Chambers into account, it seems likely that facts became eligible to be regarded as data by being provided or made available by a person or group of people who had made observations or measurements, and then imposed some form of order or structure upon them in order to facilitate analysis. Drucker's advocacy for capta as a key term and a guiding principle is rooted in the notion that the term *data* assumes a pre-existing given that can be naively recorded without being shaped by the observer. The story I have told suggests a more nuanced history of data, both as a term and a concept. There is room to embrace the word *data* and understand data to be inevitably fragmentary, imperfect, and tentative, yet simultaneously useful in their ability to organize knowledge and facilitate modes of analysis that would otherwise be impossible. In the section that follows, I will make the case for embracing concepts like situated data and data-rich literary history to convey this message, which includes potential benefits to scholarly discourse and digital humanities pedagogy.

Situating Data for the Sake of Interdisciplinarity and Digital Humanities Pedagogy

Up to now, I have attempted to show that the term *data* has a much more nuanced history that its translation to the English word "given". This revised etymology leaves room for an understanding of both the term and concept as less closely linked to naive realism than Drucker suggests. As S.I. Hayakawa once famously wrote, "the writer of a dictionary is a historian, not a lawgiver" [Hayakawa 1990, 35]. The meaning of *data* has changed over time and, at any given time, no singular meaning for the term was universal. I have pointed to one important sea change, by which data came to have special prominence in the context of empiricism. A longer essay could say much more on the subject of additional changes in prevailing notions of the term, or additional meanings that persist in specific contexts or among specialized groups. Along with the accepted range of definitions for *data*, their various connotations and their preferred rhetorical uses, have also changed over time. I think Drucker is right to point out that many today use the term *data* to suggest a kind unimpeachable concreteness rather than something partly born out of the assumptions made when selecting and organizing said data. I suspect that there is widespread disagreement among self-proclaimed empiricists about the degree to which data are "natural representations" as opposed to "situated, partial, and constitutive," but I agree that scientific rhetoric often downplays or denies the artificial, conditional, and fragmentary aspects of data [Drucker 2011].

As an alternative to avoiding the term *data*, I would argue that digital humanities should focus on challenging and complicating beliefs about the purity, objectivity, or totality of data, all the while using the word *data* frequently, and without any shame. There is a strong defense for using the word *data* mindfully and unapologetically based on the understanding that data are collected, assembled, and recorded by people (or their instruments). Data have structure, but this structure comes (at least in part) from how observations are gathered and organized, as well as shaped by the interpretive decisions made at their inception. Contextual information about these interpretive decisions is a vital component of a dataset. Many in digital humanities have already moved in this direction, shifting from notions of "good data" and "bad data" to concepts like data as situated knowledge, as described in Catherine D'Ignazio and Lauren F. Klein's *Data Feminism* (2020) or data-rich literary history as described in Katherine Bode's *A World of Fiction: Digital Collections and the Future of Literary History* (2019). The benefits of these strategies are both discursive and pedagogical.

The first benefit of this approach is to move away from an argument based on word origin. A word's origins seldom tell an accurate story of a word's historical use or contemporary valences. There are numerous Latin words with English homographs, and often their meanings have changed over time. Words such as "plastic," "stigma," "campus," "focus," "versus," "stimulus," and "sinister" all have Latin antecedents with meanings that differ greatly from their contemporary use in English. With the word *data*, there is added ambiguity because the English word "given" has more than one meaning. Describing an idea as "a given" or "a warrant" in the context of a rhetorical argument remains common, and seems to carry no implication of pre-factualness or ineligible for questioning.[20]

Capta, likewise, is Latin for taken and therefore has a certain surface appeal as a mirror term for data. However, capta can also mean "caught," "captured," and "captive." The phrase "Judæa capta" refers to the siege and capture of Jerusalem by the Romans, as well as a series of coins issued by the Roman Emperor Vespasian to celebrate these actions [Elizabeth 1845]. In the Roman Empire, the phrases *manu capta* [Sandars 1917, xlvii] and *praeda manu capta* denoted that which was "captured by hand" or "acquired by force of hands" from among Rome's enemies.[21] Further, some enslaved people ("servi") were "so called from the fact that commanders are used to *sell* their captives, and by this means to *preserve* (servare) rather than kill them" (*Justinian Institutes*, qtd. in Cameron (1972), 5).[22] These potential associations are important because they complicate the idea that data straightforwardly means *given* and capta straightforwardly means *taken* as we think of those terms today. Moreover, these examples underscore the broader point that it can be a mistake to use a word's denotation in its language of origin (or its historic denotation in English) as the primary criteria by which to judge its contemporary meaning or appeal.

Denotations, connotations, and rhetorical uses change over time, and this is good news. Such associations can change once again, and scholars of the humanities can help change them. As Helen Longino has argued, the notion of objectivity in scientific inquiry is bound up in two different senses of the term, the first emphasizing scientific realism, or the idea that science provides an "accurate description of the facts of the world as they are" [Longino 1990, 62]. The second, associated with modes of inquiry, stresses "reliance on nonarbitrary and nonsubjective criteria for developing,

accepting, and rejecting the hypotheses and theories that make up the view" [Longino 1990, 62]. Further, scientists will often "speak of the objectivity of data" but objectivity in this sense should be taken as a claim of reliability [Longino 1990, 63]. What makes data objective is "the relationship of measurements to one another within a particular dimension or kind of scale" [Longino 1990, 63]. According to Longino, it should not be assumed that these measures "are real properties of real entities" or "that their measurements provide us with an unmediated view of the natural world" [Longino 1990, 63]. In other words, there is an important difference between an etymologically inspired belief that data are natural representations of the world — given rather than taken — and the idea that scientific observation generates objective information by providing epistemological order. The two positions are based on competing definitions of objectivity. Adopting the perspective that phenomena become data only when they are structured by an investigative agent invites an important debate about the objectivity of data once they are standardized and assembled, and that is a debate that I would be glad take part in.

I would not deny the merit of distinguishing between the idea of data-as-given and capta-as-taken. However, this distinction may be more a matter of emphasis than definition. Whatever the etymological roots and contemporary associations of data, we can take capta to emphasize that the measurements, readings, observations have been taken. As I have suggested, the word *capta* may stir connotations of human bondage and violence. This association may be appropriate in some cases, especially in a moment when for-profit companies are covertly monetizing and selling data on the open market, often in direct conflict with the interests of its users. On the other hand, many data do not fit this description, and I think regarding all data as captive or hostage to their stewards is provocative, but ultimately misguided, especially where there is strong potential for unintended consequences.

<div style="text-align:right">29</div>

As previously suggested, there are matters of utility to consider when deciding between data and capta as conceptual signposts. Many in digital humanists are likely using the term *data* at least some of the time, either as a standalone key term, or in cognates or phrases like metadata, database, or data visualization. Going further down this path would allow digital humanists to compete more effectively for search engine keyword searches, which could make digital humanities scholarship more visible to audiences from the sciences. This would include direct searches, as well as related searches like "data analysis," "data analytics," and "data usage," all of which are top "related to data" searches on Google Trends. It would also encourage digital humanities to continue any previous efforts to engage directly with specialized topics such as data pedagogy, data literacy, open data, data management, and data curation.

<div style="text-align:right">30</div>

Along these lines, thinking differently about the word *data* may lead to new exchanges of information among and between disciplines. Digital humanities, as a field, has a deeply interdisciplinary history, but a precondition to such interdisciplinarity is the premise that different disciplines can learn from one another. By engaging directly with concepts related to data, we have opportunity to shifts scholarly focus from the nature of data to strategies that promote engaging critically, theoretically, and computationally with data.

<div style="text-align:right">31</div>

Catherine D'Ignazio and Lauren F. Klein take an approach of this sort in *Data Feminism* (2020). They define data as "information made tractable," which I immediately associate with Marinus' discussion of *ordinatum*. In various passages, D'Ignazio and Klein cite Donna Haraway, whose essay "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective" (1988) called for "a doctrine of embodied objectivity that accommodates paradoxical and critical feminist science projects" [Haraway 1988, 581]. Core to Haraway's "feminist objectivity" is the principle of situated knowledges. Adapting and extending Haraway's intervention, D'Ignazio and Klein argue that "one of the central tenets of feminist thinking is that all knowledge is situated" [D'Ignazio and Klein 2020, 152]. There is a more complicated discussion to be had about whether all knowledge is inherently situated or whether it is the work of feminism that situates it. Haraway seems to be arguing for the latter when she contrasts situated knowledges with "unlocatable, and so irresponsible, knowledge claims" [Haraway 1988, 583]. Either way, all data have important contexts of creation and organization, and situating them (or emphasizing them as situated) includes critical examination of those contexts.[23]

<div style="text-align:right">32</div>

D'Ignazio and Klein describe numerous practices related to examining the context of data. These practices begin with understanding the conditions of production associated with data, including information about those who constructed the data. "When approaching any new source of knowledge," they write, "it's essential to ask questions about the social, cultural, historical, institutional, and material conditions under which that knowledge was produced, as well as about the

<div style="text-align:right">33</div>

identities of the people who created it" [D'Ignazio and Klein 2020, 152]. Like Gitelman, D'Ignazio and Klein argue against the existence of anything that might be called raw data [D'Ignazio and Klein 2020, 159]. They advocate for considering the "functional limitations of the data" and "any associated ethical obligations" [D'Ignazio and Klein 2020, 153]; "exploring and analyzing what is missing from a dataset" [D'Ignazio and Klein 2020, 160]; attending to "power differentials" that have shaped and/or continue to be present in collected data [D'Ignazio and Klein 2020, 160]; and interrogating a dataset's validity — that is, the degree to which it can be said to represent the concept being analyzed [D'Ignazio and Klein 2020, 160]. They emphasize the importance of context at the stages of data acquisition, data analysis, and "framing and communication of results" [D'Ignazio and Klein 2020, 164]. Their advocacy is empowered rather than hindered by their adoption of a precise vocabulary that practitioners in a range of disciplines can recognize and quickly understand.[24]

Similarly, Katherine Bode has argued in *A World of Fiction* (2019) for an intervention in computational literary studies based on the argument that "distant reading and macroanalysis construct and seek to extract meaning from models of literary systems that are essentially deficient," which to say, "inadequate for representing the ways in which literary works existed and generated meaning in the past" [Bode 2019, 5]. Bode does not advocate that we reject data-driven literary history, as others have argued, nor does she endorse the adoption of "new, more elaborate forms of computational analysis" as a way around the conflict [Bode 2019, 6].[25] Instead, she anchors her intervention around the idea of "a scholarly edition of a literary system" — which engages explicitly with modeling practices — as "a mechanism through which to interrogate and refine conceptions of literary works and systems" [Bode 2019, 6]. As with D'Ignazio and Klein, Bode calls for increased contextualization, for "a data-rich model of a literary system is inevitably an argument shaped by not only the scholar's perception of cultural artifacts and phenomena but the complex history by which those artifacts and phenomena are transmitted to and by us in the present" [Bode 2019, 6]. Bode's use of the key phrase "data-rich" is important to her argument, in that she presents it as an alternative to labels like "distant reading," "macroanalysis," and "computational literary history" [Bode 2019, 2]. It is structurally reminiscent of "data-driven" but the adjectival "rich" has a strong contrast to the past participle "driven", both in terms of connotation and its symbolic rejection of that idea that data ostensibly occupy the driver's seat. D'Ignazio & Klein and Bode make very different interventions, but they share a common position that the term *data* has purchase, especially when that term is qualified with productively descriptive modifiers. Further, they base their work, in part, on the premise that data-driven inquiry can be executed more effectively if instilled with feminist and/or humanistic values.

If this premise is to be accepted, our conversations about how to improve upon computational inquiry can go further. Most immediately, there are implications for digital humanities pedagogy. As Ted Underwood has argued, "digital humanities classes, as currently defined, don't really teach students how to use numbers...So it's almost naive to discuss 'barriers to entry.' There is no entrance to this field" [Underwood 2018]). Underwood argues that the cultural analytics subfield, in particular, operates primarily as a social network. In such a context competency and fluency come as a result of participants' tacit knowledge and through person-to-person or small group interactions. In contrast, a well-defined curriculum with explicit pedagogies for teaching research design, data collection, data modeling, programming, and data analysis (including but not limited to quantitative analysis) would help create pathways to entry that are currently hidden for or unavailable to many people. Overt teaching of methods, Underwood argues, allows facility with methods to be more equally distributed. Discussing data literacy openly and critically strikes me as a key component of the curricular and pedagogical intervention Underwood has described.

There are wide-ranging benefits to speaking and writing about data, openly and often, with implications from the practical to the profound. As part of this strategy, we should ask our students to read about the history of the word *data*, but to include how Euclid used the word *dedoména*, how Marinus described two different potential definitions for the word, and how natural philosophers of 17th-century England, in particular, appropriated dedoména's Latin alternative *data*. We should ask them to discuss why such thinkers saw similarities between the Greek notion of data and how they were collecting and structuring information they observed and recorded. We should speak openly about crucial concepts related to data-driven inquiry, including practices aimed at contextualizing and situating data and datasets. We should press the idea that social constructions foreground and pervade what many think of as neutral or natural measures of reality, and we should openly question paradigms of thought that emphasize data as self-evident or

unproblematically objective. In short, with a sense of pride and purpose, let us say and write data, all the while remaining humanists.

Notes

[1] I would like to thank Sam Cowling, John Ladd, Rebecca Lee, and Scott Weingart for reading drafts of this essay at various stages of the writing process, as well as the peer reviewers of this piece for generously providing constructive criticism and suggestions for revision.

[2] In making this claim, Drucker gestures at two well-known positions on the relationship between "theoretical claims" and "knowledge of the world." However, these polarities can be divided into numerous subcategories, and many such positions do not fit neatly into the category of realism or constructivism. For a detailed summary of such responses, see Anjan Chakravartty, "Scientific Realism", *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Ed. Edward N. Zalta, https://plato.stanford.edu/archives/sum2017/entries/scientific-realism.

[3] In this essay, I have chosen to treat data and capta primarily as singular nouns when discussing either the term *data* or the concept of data. When referring to that which is signified by the term data, I have maintained the norm of treating data as plural. I have italicized when making statements that refer directly to key terms but not when referring to corresponding concepts.

[4] See, for example, Stan Kelly-Bootle, *The Computer Contradictionary* (MIT Press, 1995); H.E. Jensen, "Editorial Note" (1950) in H Becker, *Through Values to Social Interpretation* (Duke UP, 1952): vii-xi.

[5] At the time of writing this essay, "Humanities Approaches to Graphical Display" had 399 citations on Google Scholar, and *Graphesis* had 348. Examples of syllabi raising this topic directly or assigning Drucker on this topic include: https://cms633.github.io/assets/cms-633-syllabus.pdf; http://benschmidt.org/HDA19/syllabus__syllabus.html; https://dhcu.ca/portfolio/syllabus/; and http://www.cs.cmu.edu/~dbamman/courses/76829/schedule.html.

[6] Shawn Graham, syllabus for graduate seminar in DH, https://dhcu.ca/portfolio/syllabus/. On the original syllabus, the word *capta* is hyperlinked to Drucker's piece for *DHQ*.

[7] It is conceivable that DH would begin publishing papers on captabases and metacapta, but many who would prefer the term *capta* are probably using data cognates and phrases out of convenience or preference. Kelly-Bootle satirically raises the possibility of terms like capta entry, capta processing, and captabase in *The Computer Contradictionary* (1995) [Kelly-Bootle 1995, 52]. When I attempt a Google search for metacapta, I get the response, "Did you mean metacarta, metacipta, metacapital, metacafe?" On Google Scholar, I get results for metacaptcha and metacapita. (Results collected in Summer 2020.)

[8] The *OED* cites the earliest use of "mathematics" as "(a collective term for) geometry, arithmetic, and certain physical sciences involving geometrical reasoning, such as astronomy and optics" to Christopher Langton, circa 1545 [OED 2020d] For more on the unification of geometrical and numerical thought, see Patrick Suppes et. al., *Foundations of Measurement Volume 2*: *Geometrical, Threshold, and Probabilistic Representations* (Elsevier, 2014): 80-81.

[9] English language translations of Euclid's *Elements of Geometry* date as far back as 1570. See Adam Clarke, *The Bibliographical Miscellany* (London: W. Baynes, 1806): 77-78.

[10] The *OED* credits William Batten's *A MOST PLAINE and easie way for the finding of the Sunnes Amplitude and Azimuth, and thereby the Variation of the Compasse, by Logarithme* (1630) under a truncated title. Note, also, that the word *data* appears in keyword searches of databases like Early English Book Online (EEBO) before these 1630 because editions in Latin, or English language editions with Latin quotations or epigraphs, were quite common. In Rosenberg's words, "the Latin word *data*, as a conjugation of the verb 'dare', was in constant use during the seventeenth and eighteenth centuries...but *data* in Latin rarely translates to 'data' in English" [Rosenberg 2013, 20].

[11] The *OED* credits T. Urquhart's *Trissotetras* (1645) as the earliest known example of "data" as a count noun, and credits Robert Morden's *An Introduction to Astronomy, Geography, Navigation, and other mathematical sciences, made easie by the description and uses of the cœlestial and terrestial globes* (1702) as the earliest known example of data as a mass noun.

[12] See Michael Frede, "Stoic vs. Aristotelian Syllogistic." In *Essays in Ancient Philosophy* (U of Minnesota P, 1987), pp. 99–124; John Spangler Kieffer, *Galen's Institutio Logica: English Translation, Introduction, and Commentary* (JHU Press, 2020); and Lucas Francisco Mateo Seco and Giulio Maspero. *The Brill Dictionary of Gregory of Nyssa* (Brill, 2010).

[13] See David Dunér and Christer Ahlberger, *Cognitive History: Mind, Space, and Time* (Walter de Gruyter GmbH & Co KG, 2019), p. 100.

[14] The *OED*'s etymological entry for "phenomenon" describes, "things that appear, appearances, phenomena, celestial phenomena (title of a work by Eudoxus, versified by Aratus), use as noun of neuter of 'φαινόμενος' appearing, apparent (to the senses or mind), passive present participle of 'φαίνειν' to show, cause to appear" [OED 2020d]. For more on Eratosthenes' *Geography*, see Eratosthenes, *Eratosthenes' "Geography"* (Princeton University Press, 2010).

[15] See Hadley Wickham, "Tidy Data," *The Journal of Statistical Software*, vol. 59 (August 2014), pp. 1-23. In contrast to an observation, values are specific numbers or string, and a variable "contains all values that measure the same underlying attribute (like height, temperature, duration) across units" [Wickham 2014, 3].

[16] See Peter H. Reill, *Vitalizing Nature in the Enlightenment* (University of California Press, 2005); Lorraine Daston and Peter Galison, *Objectivity* (Zone Books, 2010); Edward Dolnick, *The Clockwork Universe: Isaac Newton, The Royal Society, and The Birth of the Modern World* (Harper Collins, 2011).

[17] For more on early modern associations between knowledge and sin, as well as associations between epistemological errors and sin, see Peter Harrison, *The Fall of Man and the Foundations of Science* (Cambridge University Press, 2007).

[18] Taisbak translates πορίσαοθαι (*porizesthai*) cognate to the verb "to provide," as "available" in this context [Taisbak 2003, 24].

[19] Taisbak adds that "porimon" is "vital to the interpretation of the *Data*," but Euclid "uses the word as being self-explanatory. It seems to be imported from everyday market prose meaning 'procure,' 'furnish,' 'get (for oneself):' any kind of purchase without specification of method and currency" [Taisbak 2003, 242].

[20] The word "plastic" comes from the Latin for "capable of shaping or molding" [Sullivan 1860, 165] while "stigma" conveys the "mark of a pointed instrument, puncture, tattoo-mark, or brand" [Sullivan 1860, 206]. "Campus" is Latin for "field," "focus" is Latin for "hearth," and "versus" is Latin for a line of poetry [Sullivan 1860, 194, 103, 220]S. A "stimulus" was a prod or a goad, which is to say a sharp stick used to urge livestock forward (Vaan 587). "Sinister" comes from the Latin word for "left," and eventually came to mean evil or unlucky, but was purportedly used by the Romans "in the opposite sense," such that am "avis sinistra" or "bird in the left hand...was esteemed a happy omen" [Encyclopaedia Perthensis 1816, 7]. If the frame is widened to include linguistic doublets and cognates, the list balloons to the point of absurdity. The words "dose," "antidote," "anecdote," "date," "donatio," "addendum," "addition," "edition," "editor," "mandate," and "mandatory" all share the Greek or Latin root word for the verb "to give," but they have a vast range of meanings and associations today [Shipley 2001, 73]. Perhaps most ironically, the portmanteau "anecdata" replaces part of the Greek root word for "given" with the Latin word for "given." See Anu Garg, "A.Word.A.Day" (http://www.wordsmith.org/words/anecdata.html) for a summary of the term "anecdata" and Shipley for a summary of how "anecdote" originates with the Greek *didonai* [Shipley 2001, 73].

[21] Such phrases referred especially "properties taken in war" and could describe compensation paid by "the vanquished...either in kind (*vectigal*), with money (*stipendium*), or with a tribute (*tributum*)" [Vico 2000, 492].

[22] They are also said to be *mancipia*, because they were taken by hand (*manu capta*) from the enemy (*Justinian Institutes*, qtd. in Cameron (1972), 5).

[23] See also Jill Walker Rettberg, "Situated Data Analysis: A New Method for Analysing Encoded Power Relationships in Social Media Platforms and Apps," *Humanities and Social Sciences Communications* 7, no. 1 (June 17, 2020): 1–13. https://doi.org/10.1057/s41599-020-0495-3. Rettberg also frames "situated data analysis" as an extension of Haraway's concept of "situated knowledge."

[24] For a further indication of such terminological compatibility, see Timnit Gebru et. al. "Datasheets for Datasets," *ArXiv:1803.09010 [Cs]*, March 19, 2020. http://arxiv.org/abs/1803.09010. The authors note, "Despite the importance of data to machine learning, there is no standardized process for documenting machine learning datasets" [Gebru et al. 2020, 1]. They propose a standard to "encourage careful reflection on the process of creating, distributing, and maintaining a dataset, including any underlying assumptions, potential risks or harms, and implications of use" and to ensure that "dataset consumers...have the information they need to make informed decisions about using a dataset" [Gebru et al. 2020, 2].

[25] Bode does not eschew methodological rigor altogether. She writes, "Although it is a premise of the book that data- rich literary history should not focus on the methods of analysis used to the detriment of the object analyzed, in a field that attempts to understand literature and culture by applying techniques devised for other purposes a critical approach to methodology is essential" [Bode 2019, 9]. Her critique of methodological emphasis lies in a tendency of others to rely on "trialing the newest or most innovative digital methods" instead of those best suited to the subject matter and "to responding to the requirements of humanities inquiry" [Bode 2019, 9].

Works Cited

Austin 1934 Austin, Roland G. "Zeno's Game of Τάβλη (A. P. ix. 482)," *The Journal of Hellenic Studies* 54 (1934): 202–5. https://doi.org/10.2307/626864.

Bode 2019 Bode, Katherine. *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor: U of Michigan P, 2019.

Campbell-Kelly 2003 Campbell-Kelly, Martin, et al., *The History of Mathematical Tables: From Sumer to Spreadsheets*, Oxford, Miss.: OUP, 2003.

Chambers 1728 Chambers, Ephraim. *Cyclopædia, or, An Universal Dictionary of Arts and Sciences* (1728). ARTFL Reference Collection. https://artfl-project.uchicago.edu/content/chambers-cyclopaedia.

D'Ignazio and Klein 2020 D'Ignazio, Catherine, and Lauren F. Klein, *Data Feminism*. Cambridge, MIT Press, 2020.

Drucker 2011 Drucker, Johanna. "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5.1 (2011): n.p. http://www.digitalhumanities.org//dhq/vol/5/1/000091/000091.html.

Drucker 2014 ——. *Graphesis: Visual Forms of Knowledge Production*. Cambridge: Harvard UP, 2014.

Elizabeth 1845 Elizabeth, Charlotte. *Judæa Capta: An Historical Sketch of the Destruction of Jerusalem by the Romans*. London: W.H. Dalton, 1845.

Encyclopaedia Perthensis 1816 *Encyclopaedia Perthensis, Or, Universal Dictionary of the Arts, Sciences, Literature, Etc.: Intended to Supersede the Use of Other Books of Reference*. Edinburgh: J. Brown, 1816:

Evans and Berggren 2018 Evans, James, and J. Lennart Berggren, *Geminos's Introduction to the Phenomena: A Translation and Study of a Hellenistic Survey of Astronomy*. Princeton: Princeton UP, 2018.

Furner 2016 Furner, Jonathan. "'Data': the data," *Information Cultures in the Digital Age.* Eds. M. Kelly, J. Bielby, Wiesbaden: Springer, 2016. 287-306. https://doi.org/10.1007/978-3-658-14681-8_17.

Gebru et al. 2020 Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. "Datasheets for Datasets." ArXiv:1803.09010 [Cs], March 19, 2020. http://arxiv.org/abs/1803.09010.

Gitelman 2013 Gitelman, Lisa. "Introduction," *"Raw Data" is an Oxymoron*, ed. by Lisa Gitelman. Cambridge: MIT Press, 2013. 1-14.

Haraway 1988 Haraway, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies* 14, no. 3 (1988): 575–99, https://doi.org/10.2307/3178066.

Hayakawa 1990 Hayakawa, Samuel Ichiyé. *Language in Thought and Action*. New York: Houghton Mifflin Harcourt, 1990.

Heath 1921 Heath, Sir Thomas. *A History of Greek Mathematics Volume 1: From Thales to Euclid*. Oxford: Clarendon Press, 1921.

Justinian 1972 *Justinian Institutes*, Lib. 1, Tit. 3., "De Jure Personarum." Qtd. in Archibald Alexander Cameron, *Protestantism and Its Relation to the Moral, Intellectual and Spiritual Developments of Modern Times*. Ottawa: Joseph Loveday, 1972.

Kelly-Bootle 1995 Kelly-Bootle, Stan. T*he Computer Contradictionary*. Cambridge: MIT Press, 1995.

Kitchin 2014 Kitchin, Rob. "1. What Are Data?" section on "Conceptualizing Data," *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage Publications e-book edition, 2014.

Longino 1990 Longino, Helen E. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton UP, 1990.

Marinus 1661 Marinus, "Euclides Data, a Commentary or Preface Written by the Phylosopher Marinus on Euclides Data," in *Euclid's Elements of Geometry*. London: John Leeke & George Serle, 1661.

OED 2020a "almanac, n." *OED Online*. (Oxford University Press, June 2020), accessed October 16, 2020. https://www.oed.com/view/Entry/5564.

OED 2020b "ephemeris, n." *OED Online*. (Oxford University Press, June 2020), accessed October 16, 2020. https://www.oed.com/view/Entry/63209.

OED 2020c "data, n." *OED Online*. (Oxford University Press, June 2020), accessed July 01, 2020. https://www.oed.com/view/Entry/296948.

OED 2020d "mathematics, n." *OED Online*. (Oxford University Press, June 2020), accessed October 15, 2020. https://www.oed.com/view/Entry/114974

OED 2020e "phenomenon, n." *OED Online*. (Oxford University Press, June 2020), accessed October 15, 2020. https://www.oed.com/view/Entry/142352.

OED 2020f "table, n." *OED Online*. (Oxford University Press, June 2020), accessed October 25, 2020. https://www.oed.com/view/Entry/196785.

Rosenberg 2013 Rosenberg, Daniel. "Data Before the Fact," *"Raw Data" is an Oxymoron,* ed. by Lisa Gitelman (Cambridge: MIT Press, 2013): 15-40.

Sandars 1917 Sandars, Thomas Collett. *The Institutes of Justinian*. New York: Longmans, 1917.

Seco et al. 2010 Seco, Lucas, Francisco Mateo, and Giulio Maspero. *The Brill Dictionary of Gregory of Nyssa*. BRILL, 2010.

Shipley 2001 Shipley, Joseph Twadell. *The Origins of English Words: A Discursive Dictionary of Indo-European Roots*. Baltimore: Johns Hopkins UP, 2001.

Sullivan 1860 Sullivan, Robert. *A Dictionary of Derivations; Or, An Introduction to Etymology: On a New Plan.* Dublin: Marcus and John Sullivan, 1860.

Taisbak 2003 Taisbak, Christian Marinus. *Dedomena: Euclid's Data or The Importance of Being Given*, The Greek Text translated and explained by Christian Maris Taisbak. Copenhagen: Museum Tusculanum Press, University of Copenhagen, 2003.

Underwood 2018 Underwood, Ted. "A Broader Purpose," Varieties of Digital Humanities Panel, MLA, Jan 5, 2018. https://tedunderwood.com/2018/01/04/a-broader-purpose/.

Vaan 2008 Vaan, Michiel de. *Etymological Dictionary of Latin and the Other Italic Languages*. 2008. Leiden, Boston: Brill, 2018.

Vico 2000 Vico, Giambattista. *Universal Right*. Trans. and ed. by Georgio Pinton & Margaret Diehl. Amsterdam: Ridopi, 2000.

Wickham 2014 Wickham, Hadley. "Tidy Data." *The Journal of Statistical Software* 59 (August 2014): 1-23, https://doi.org/10.18637/jss.v059.i10.