"Data": The data

Jonathan Furner

Abstract

While many scholars in information science have understandably focused on the concept of "information" as foundational, some authors have identified other concepts as having similarly foundational status. Two that are regularly suggested as candidates are "data" and "document." Oddly, perhaps, for such a basic term, "data" has not been as frequently subject to probing analysis in the scholarly literature as "information"; and although "document" has long been a term of special interest to historians of the European documentation movement, some of whom continue to develop a document theory, there is little consensus on the precise nature of the conceptual relationship between "data" and "document." In this paper, a review is conducted of historical interpretations of "data," and relationships with contemporary conceptions of "document" are explored. The conclusion is reached that, current practice notwithstanding, it is not in fact the case that documents are made up of data, nor that the document is a species of dataset: rather it is the other way round, in both respects. A dataset is made up of documents; and the dataset is a species of document.

> "For a science like information science (IS), it is of course important how fundamental terms are defined."
> (Capurro & Hjørland, 2003, p. 344)

Since the emergence of the field known as information science in the 1950s, scholars have understandably focused on the concept of "information" as foundational. With their *ARIST* chapter of 2003, for example, Rafael Capurro and Birger Hjørland's goal was "to review the status of the concept of information in IS" (Capurro & Hjørland, 2003, p. 344). Notwithstanding this preoccupation with "information," some authors have identified other concepts as having similarly foundational status. Two that are regularly suggested as candidates are "data" and "document."

Oddly, perhaps, for such a basic term—and as commentators have repeatedly remarked (see, for example, Borgman, 2007)—"data" has not been as frequently subject to probing analysis in the scholarly literature as "information."[1] Given that "data" is now, in 2015, somewhat of a "word of the moment," used in such recently prevalent terms as big data, linked data, open data, data governance, data infrastructure, data mining, data protection, data quality, data science, data visualization, and data wrangling, a review of conceptions of "data" along the same lines as Capurro and Hjørland's treatment of "information" seems long overdue. In the meantime, we can benefit from those comparative studies that situate "data" in relation to cognate terms like "information," "knowledge," and "wisdom" (see, for example, Zeleny, 1987; Ackoff, 1989; Thow-Yick, 1994; Rowley, 2007; Frické, 2009),[2] as well as from explorations of the cultural, disciplinary, and historical contexts for data production and consumption (see, for example, Gitelman, 2013; Markham, 2013), and expositions of the logic of data modelling practices (see, for example, Kent, 1978; Simsion, 2007).

"Document" has long been a term of special interest to historians of the European documentation movement, some of whom continue to develop a document theory based on the ideas of the Belgian lawyer Paul Otlet (1868–1944) and the French librarian Suzanne Briet (1894–1989), among others (see, for example, Buckland, 1997; Day, 2001; Frohmann, 2004; Lund, 2009). More recently, a renewed interest by philosophers, cultural historians, and anthropologists addresses the notions of document, documentality, and document act as foundational concepts (see, for example, Riles, 2006; Ferraris, 2013; Gitelman, 2014; Smith, 2014).

There is little consensus on the precise nature of the conceptual relationship between "data" and "document." The default position appears to be the view that all documents are in some sense made up of data. Sometimes, it is allowed that there are several different types of data, and one of those in particular is the kind of data that makes up documents. A document is a special kind of aggregation of data: that is, it is a species of dataset. In this case, "data" is the primary concept: if data did not exist, documents could not; even though data do exist, documents need not.

The position I wish to develop in this paper, however, is that it is not in fact the case that documents are made up of data. On the contrary, it is the other way round: datasets are made up of documents. There are several different types of document, and one of those in particular is the kind of document that makes up datasets. And, given that aggregations of documents can themselves be considered as documents, a dataset is a species of document. In this case, "document" is the primary concept: if documents did not exist, data could not; even though documents do exist, data need not.

---

1 For the purposes of a previous paper (Furner, 2014), I found more than fifteen different scholarly publications sharing the title "What is information?" A similar search for works titled "What is [*or* are] data?" turned up two sources, neither of particular utility.

2 It is entirely possible that not all of these scholarly interventions were inspired by T. S. Eliot's *The Rock*: "Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?" (Eliot, 1934, p. 7).

In this paper, I want to try to show why I believe that "document" is the primary concept, and why I believe that it is important to take that position over others. There are two main parts to the paper. In Section 2, I review some of the analyses that have been undertaken of the concept of "data," and give a historical account of the uses to which the word "data" has been put since its coinage in classical Latin. In Section 3, I briefly present the well-known way of thinking about the nature of documents previously articulated by Michael Buckland and others, and show how (what I take to be) the most useful interpretation of "data" may be incorporated into Buckland's framework, with a result that provides corroboration for my conclusion in Section 4.

1 The Nine Lives of "Data"

Analyses of the concept of "data" proceed in a variety of ways. We may distinguish among analyses that take an extensional approach (i.e., more or less exhaustively listing the things, or kinds of things, that fall under the generic heading of "data"); those that take an intensional approach (i.e., identifying the properties that something must have if it is to be treated as data); and those that take a classificatory approach (i.e., recognizing that an individual concept like "data" may have, have had, or could have multiple senses, and that these senses may be categorized in accordance with similarities in function, context, etc.).[3] We may also highlight historical approaches, which instead of or in addition to conducting logical and/or computational analyses of the necessary properties of concepts, allow authors to consider the culturally specific development of the meanings of terms like data over time (see, for example, Rosenberg, 2013).

Floridi, for example, uses the third of these methods to identify four "principal interpretations" of the meaning of "data": epistemic, informational, computational, and diaphoric.[4] The epistemic interpretation equates data with "the basic *assumptions* or empirical *evidence*" on which reasoning can be based; the informational interpretation equates data with information, "in the ordinary sense in which information is equivalent to some content … about a referent" (2008, p. 234); the computational interpretation sees data as "collections … of *binary elements*" (p. 235); while the diaphoric interpretation defines data as distinctions, differences, or "lacks [sic] of uniformity" (p. 235). Floridi submits that the diaphoric interpretation is "the most fundamental and satisfactory" (p. 234), and further suggests that it may be applied at three separate levels; he distinguishes among the kinds of difference that may be found at the level of the world (diaphora *de re*), among the physical states or signals of a system (diaphora *de signo*), and among the symbols of a code (dia-

3 Somewhat similarly, Christine Borgman (2014) distinguishes among "definitions by example," "operational definitions," and "categorical definitions."

4 Floridi notes that "the word 'data' comes from the Latin translation of a work by Euclid entitled *Dedomena*." This is not strictly true, of course: the word "data" was in general use in Latin prior to the translation of Euclid's work.

phora *de dicto*). Floridi states three "principle[s] of … neutrality" that are satisfied by the diaphoric interpretation (p. 236): data are not substantive entities but rather *relata* (that is, external properties); data may be typed according to "the sort of analysis conducted and … the level of abstraction adopted" (as, for example, primary data, secondary data, metadata, operational data, and/or derivative data; p. 236); and data can be meaningful "*independently* of an informee" (p. 236).

To Floridi's taxonomy of four, we might add at least five other families of interpretation, for a total of nine, beginning with the classical interpretation that dates from antiquity.

1.1 The Classical Interpretation: Data as Gifts (ca. 100 BCE–)

In classical Latin (the form of the language used in the period roughly between 100 BCE and 200 CE), the present active infinitive of a first-conjugation verb has the suffix *-are*; examples are *amare* ("to love"), and *dare* ("to give")—otherwise commonly denoted by the first person singular of the verb's present active indicative, i.e., *amō* ("I love"), and *dō* ("I give").[5] The nominative masculine singular form of the perfect participle of *dō* is *datus* ("given"); *datum* is this participle's nominative neuter singular form, while *data* is both the nominative neuter plural and the nominative feminine singular.

The perfect participle—which is always passive—functions in the same way in Latin as an adjective in English: e.g., *Pecunia data*, "The having-been-given money," or "The money that has been given," or simply "The money given." Used as a substantive, *datum* can mean "gift" (in the literal sense of "that which is given");[6] the plural *data*, then, may be translated as "things given" or "gifts."[7]

During the Roman Empire's period of decline, written Latin also suffered widespread debasement. The medieval Latin (ca. 600–1500 CE) that ultimately supplanted the Late Latin dialect (ca. 200–600 CE) was more hospitable to the introduction of words, meanings, and grammatical structures from various vernaculars, and the Latin dictionaries produced by scholars across Europe during this time reflected such change. The *Summa Grammaticalis*

---

5 *Dō* itself has been hypothesized to descend ultimately from the reconstructed Proto-Indo-European root $*deh_3$-. The origins of the English words *give* and *gift*, on the other hand, have been tentatively traced to the reconstructed Proto-Indo-European root $*g^{h}h_1b^{h}$-, from which source also come not only the German *geben* ("to give") but also, perhaps somewhat counter-intuitively, the Latin *habeō* ("I have, hold").

6 A Latin synonym of *datum* is the second-declension neuter noun *dōnum* (plural *dōna*), from which is derived the first-conjugation verb *dōnō* ("I give"), with its present active infinitive *dōnāre* ("to give") and perfect passive participle *dōnātus* ("given"), and the third-declension feminine noun *dōnātiō* (plural *dōnātiōnēs*). The links to the English verb "donate" and noun "donation" are clear.

7 The *Oxford Latin Dictionary* (2nd ed.; *OLD2*; Glare 2012) cites usages of the plural *data* in this way in the comedy *Asinaria* by Titus Maccius Plautus (ca. 254–184 BCE), the poetry of Gaius Valerius Catullus (ca. 84–54 BCE), the elegies of Sextus Propertius (ca. 50–15 BCE), and the dialogues of the Stoic philosopher Lucius Annaeus Seneca (ca. 4 BCE – 65 CE).

Quae Vocatur Catholicon ("The Grammatical Summary That is Called *Catholicon*"),[8] compiled in the late thirteenth century by the Dominican scholar Giovanni Balbi (d. 1298)[9] of the republic of Genoa on the northwestern Italian coast, was one of the most widely used Latin dictionaries in medieval Europe, and in fact was one of the first books to be printed (at Mainz in 1460, possibly by Gutenberg; see Balbus, 1971). It contains an entry for *dō*, of course, with several of its conjugated forms listed including *datus*, *datum*, and *data*, but none of these appears as a separately defined headword.

The first clear English-language definition of either *datum* or *data* to appear in a Latin–English dictionary seems to be the one given more than a century later in Thomas Thomas's *Dictionarium Linguae Latinae et Anglicanae* of 1587 (Thomas, 1972). Here, the Latin headword *dătŭm* is defined as "A thing given, a gift delivered or sent."

1.2 The Documentary Interpretation: Data as Metadata (ca. 100 BCE–)

The medieval practice of glossing—readers' handwritten annotation of manuscripts, often including Middle English translations of the Latin words and phrases found there—generated a new genre in the fifteenth century: the compilation by scholars of glosses from various sources, to form lists of Latin words with accompanying Middle English equivalents.[10] The provenance of the various surviving manuscripts of these bilingual glossaries is difficult to untangle, but two works that have circulated under different titles display great similarities, and it has been demonstrated that the authors of both are likely to have relied on the *Catholicon* and other readily accessible Latin dictionaries for their Latin vocabularies. *Medulla Grammatice* ("The core of grammar") is one of the first such glossaries to have been completed, perhaps as early as 1425;[11] *Hortus* [or *Ortus*] *Vocabulorum* ("The garden of words") dates from roughly the same period.

Surviving versions of *Medulla Grammatice* list *dō* with the Middle English equivalents of *to gyve*, *to ȝevyn*, and *to geve*—this was an era both of unstandardized spelling and a general transition from the use of the letter yogh (*ȝ*) to (depending on the context) *g*, *gh*, *w*, or *y*—and some list *dōnō* with the same meaning.[12] *Dōnum* is usually listed with the meaning *a gyft*, and *datius* (or *datiuus*), *-a*, *-um* with the meaning *gevyng*, but not *datus*, *-a*, *-um* as such. The first printed version of *Hortus*, on the other hand, supplies possibly the earliest appearance of *data* as a separate headword in a Latin–English glossary—albeit with an entry itself written entirely in Medieval Latin, which may be translated roughly

8 *Catholicon* derives from the Greek *Καθολικόν*, "Universal."

9 Giovanni Balbi was also known as Johannes Balbus, Johannes de Janua, and Johannes Januensis, among other variants.

10 Middle English is a name for the family of dialects spoken (and written) in England between ca. 1100 and 1500 CE. It was preceded by Old English (a.k.a. Anglo-Saxon; ca. 500–1100 CE), and succeeded by Early Modern English (ca. 1500–1650 CE).

11 Sometimes *grammatice* has been written as *grammaticae* or *grammatices*.

12 See, for example, Van Zandt-McCleary (1958), Tremblay (1968), and Huntsman (1973).

as follows: "*Data* or *datum* is a mark written at the end of ordinances or letters, by which means it can be determined on what month or day each letter was published."[13] There is some evidence to suggest that the form of *data* being referenced here is actually the feminine singular, rather than the neuter plural: the *Oxford English Dictionary* (3rd ed.; *OED3*; Simpson, 2000), for example, notes in its entry for the English noun "date" that the feminine singular *data* was used as a substantive in both classical and post-classical Latin to mean "date," in the specific sense of a statement at the beginning or end of a letter—a statement not only of the time at which, but usually also of the place from which, the letter was sent.[14] From this, we may infer that, rather remarkably, one of the earliest uses of *data* by English writers was to refer to what we would now likely recognize as a piece of *meta*data.

1.3 The Ecclesiastical Interpretation: Data as Gifts of God (ca. 1614–)

Perhaps unsurprisingly, given the primary concerns of the age, the first use of *data* as a word in an otherwise English-language text appears to have been in the clergyman Thomas Tuke's *Nevv Essayes* of 1614. Tuke (ca. 1580–1657) writes: "Every Sacrament is a Mysterie, but every Mysterie is not a Sacrament. Sacraments, are not *Nata*, but *Data*: Not Naturall, but by Divine appointment …" (Tuke, 1614, pp. 70–71). Searches of EEBO-TCP[15] reveal the use of *data* as an element in several Latin phrases that occasionally appear in religious texts of early seventeenth-century England—for example, *gratia gratis data* ("grace freely given"), and *data desuper* ("given from above"). More often than not, of course, the writer assumes that the reader will understand that the unnamed giver in these scenarios is God.

By the middle of the seventeenth century, it appears, the sense of *data* as (specifically) "that which is given by God" is well established. The poet John Donne (1572–1631) uses *data* in this way in his sermon of November 19, 1627, transcribed for publication in 1649

13 See *Hortus Vocabulorum, 1500* (1968): "Data vel datum est characterizatio in fine constitutionum [i.e., constitutiarum?] vel litterarum inscripta. ex qua cognosci potest quo mense vel die instrumentū[m] vel quelibet littera emanauit. f.p."

14 This sense of "date" in modern English is defined by the *OED3* as "A statement in a document, letter, book, or inscription, of the time (and often place) of enactment, writing, publication, manufacture, etc." The *OED3* gives several examples of the Latin use of the feminine singular *data* in the *Epistulae ad Atticum* of Marcus Tullius Cicero (106–43 BCE); others may be found by searching *Classical Latin Texts* at http://latin.packhum.org/. The *Dictionary of Medieval Latin From British Sources*, Vol. 1: A–L, Fasc. 3 (Latham, 1986) gives "date or dating clause" as sense 13b of *dare*, citing a number of sources from the period 1095–1404. The general use of *dare* to mean "to issue a formal letter or other document" (13a) is recorded in works from the period 601–1398, including the *Historia Ecclesiastica Gentis Anglorum* by the Northumbrian monk Bede.

15 The Text Creation Partnership [TCP] "creates standardized, accurate XML/SGML encoded electronic text editions of early print books … transcrib[ing] and mark[ing] up the text from the millions of page images in ProQuest's *Early English Books Online* [EEBO] …"; see http://www.textcreationpartnership.org/home/.

as one of his *Fifty Sermons* (Donne, 1649, p. 132), as does the clergyman Edward Boughen (1587–1660) in his *Master Geree's* Case of Conscience *Sifted* (Boughen, 1650, p. 11).

1.4 The Geometric Interpretation: Data as Geometric Premises (ca. 1645–)

"Euclid" is the anglicized form of the ancient Greek name Εὐκλείδης (Eukleidēs) used by the celebrated mathematician who flourished in Alexandria, Egypt, around 300 BCE. Euclid's sequence of books known as *Στοιχεῖα* (*Stoicheia*; "Elements") is an unrivalled candidate for the title of the most influential mathematical work of all time; it established the principal axioms and theorems of the mathematical system that came to be called Euclidean geometry. The *Elements* was first translated from the original Greek into Arabic in the eighth century, and then from Arabic into Latin in the twelfth;[16] the first printed edition (of a Latin translation) appeared in 1482, and a first English version, translated by Henry Billingsley, was printed in 1570.

Other works by Euclid that have survived include *Δεδομένα*, *Dedoména*—so-called because it opens with the neuter plural form of the passive perfect participle of *δίδωμι* (*dídōmi*, "I give")—which deals with the nature of that-which-is-given in geometrical problems, and ways of deducing what additional facts are also "given" (i.e., determinable) once the known facts, the premises, are taken into account. When, in the twelfth century, *Dedoména* was translated into Latin directly from the Greek, it was given the title *Data*. The first English translation of *Data*, by two "students of mathematics" (John Leeke and George Serle), was included with an edition of *Elements* in 1661 (Euclid, 1661); the work would have been known to some English writers before then, of course, through the various printed editions that existed of its Latin translation, but the relatively late appearance of its English translation indicates that Euclid's *Data* was not widely disseminated in England before the second half of the seventeenth century.

The earliest use of "data" in an English-language work on geometry appears to be in *The Trissotetras, or, A Most Exquisite Table for Resolving All Manner of Triangles* of 1645. Here, the Scottish maverick Thomas Urquhart (1611–60) includes the following definition of "data" in a "Lexicidion" (Urquhart, 1645, p. 94): "[T]he parts of a Triangle, which are given us, whether they be Sides or Angles, or both." The English hydrographer Joseph Moxon (1627–1691) was the first to use "data" as a headword in a specialized English-language dictionary (his *Mathematicks made easie* of 1679), with this definition: "A Term in Geometry for something proposed or known, in order to the finding out of other things unknown. As two Sides and an Angle given in a Triangle, to find the third Side; Here two Sides and an Angle are the Data. See *Euclid's Data*." (Moxon, 1679, p. 40).

16 Latin translations may well have been available in both classical and post-classical periods, but no copies of these have been preserved.

1.5 The Mathematical Interpretation: Data as Mathematical Premises (ca. 1704–)

By the start of the eighteenth century, the geometric interpretation of "data" had broadened to include any given mathematical knowns, regardless of the area of application. In John Harris's *Lexicon technicum: or, An universal English dictionary of arts and sciences* of 1704, "data" is defined as "the Term in Mathematicks for such things or quantities as are supposed to be given or known: in order to find out thereby other things or quantities which are unknown or sought for, and *Euclide* uses the Word *Data*, (of which he hath a particular Tract) for such Spaces, Lines, and Angles, as are given in Magnitude, or to which we can assign others equal. See *Given*." (Harris, 1704). Harris's definition is echoed in Nathan Bailey's *An universal etymological English dictionary* of 1721 ("Such Things or Quantities as are supposed to be given or known, in order thereby to find out Things or Quantities which are unknown, and sought for"; Bailey, 1721); Benjamin Martin's *Lingua Britannica reformata* of 1749 ("a term in mathematics for such things or quantities as are known or given, in order to find out other things thereby, which are unknown"; Martin, 1749); and the first edition of the *Encyclopædia Britannica* of 1771 ("among mathematicians, a term for such things or quantities as are given or known, in order to find other things thereby that are unknown. Euclid uses the word data [of which he hath a particular tract] for such spaces, lines, and angles as are given in magnitude, or to which we can assign others equal"; *Encyclopædia Britannica*, 1771).

1.6 The Epistemic Interpretation: Data as Evidence (ca. 1648–)

From the newly expanded sense of "data" as "any mathematical premises," the step to "any premises" *tout court* was short, and seems to have been decisively taken in the early eighteenth century (although Rosenberg [2013, pp. 20, 38] cites a 1648 work, *A brief vindication of three passages in the* Practical catechisme [Hammond, 1648], by the clergyman Henry Hammond [ca. 1605–60] as an even earlier source of "data" used in this sense[17]). In 1728, Ephraim Chambers included a generalized definition in his *Cyclopædia: or, An universal dictionary of arts and sciences* (Chambers, 1728):

> *Things given*, a Term used in Mathematicks, Philosophy, &c. implying certain Things, or Quantities supposed to be given, or known, in Order, from them, to find out other Things or Quantities, which are unknown, or sought for. A Problem or Question generally consists of two Parts; *Data* and *Quæsita*. See PROBLEM, &c.*Euclid* has an express Treatise of *Data*;

17 "… [W]ere, I say, all this granted to you, yet sure from all this heape of *data* (if they were *concessa* too) it would not follow that it was necessary, or so much as tolerably well done, to abolish all set Formes in the publique service of God, which was the prime thing by that *View* insisted on." The original source of this quotation is actually a letter of November 4, 1646, from Hammond to Francis Cheynell (1608–65), published in 1647 in *A copy of some papers past at Oxford, betwixt the author of the* Practicall cathechisme*, and Mr. Ch.* (Hammond, 1647, pp. 83–84).

> Wherein he uses the Word for such Spaces, Lines and Angeles as are given in Magnitude; or to which we can assign others equal. See GIVEN. 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; called also the *given* Quantity, Number, or Power: and such Things as are known; from whence either in Natural Philosophy, the animal Mechanism, or the Operation of Medicines, we come to the Knowledge of Things before unknown, are now frequently in Physical Writers call'd *Data*.

While there is no entry for "data" in Samuel Johnson's *A dictionary of the English language* (Johnson, 1755), Noah Webster's *An American dictionary of the English language* (Webster, 1828) paraphrases Chambers: "Things given, or admitted; quantities, principles or facts given, known, or admitted, by which to find things or results unknown." Joseph Worcester follows suit in his *Dictionary of the English language* (Worcester, 1860)—"Truths or premises given or admitted, from which to deduce conclusions; the facts from which an inference is drawn"—as does Charles Annandale in his revised edition of John Ogilvie's *The imperial dictionary of the English language* (Ogilvie, 1882)—"Something given or admitted; any condition, quantity, or other mathematical premiss, which is given in a particular problem." For two centuries, and at least until the late 1800s, the epistemic interpretation of "data" was the primary one in English.

1.7 The Informational Interpretation: Data as Attribute-Values (ca. 1630–)

A major shift in the dominant interpretation of the concept of *data* began to take place in the second half of the nineteenth century. Accompanying the rapid development of the statistical and social sciences came the proliferation of systematically organized tables of numerical values, recording and reporting the frequencies and quantities resulting from observations and measurements conducted in accordance with the principles and standards of scientific method. The contents of these tables—the "givens" that, once collected and organized, became the raw materials for new, sophisticated forms of quantitative analysis—began to be known as data. In this way, the notion of data as "content … about a referent" pre-dates by some distance the use in computer science, from the 1960s onwards, of the term "database" to talk about structured collections of recorded instances of (typically, but not always, numeric) attribute-values.[18]

---

18 Even less widely appreciated is the fact that examples of such tables with captions including the word "data" may be found in publications dating from the 1600s onwards. The *OED3* (Simpson 2000–) lists the source of the first appearance of "data" playing this role as p. 3 of a pamphlet of 1630, *A most plaine and easie way for the finding of the sunnes amplitude and azimuth, and thereby the variation of the compasse, by logarithme* (Batten, 1630), written by Sir William Batten (ca. 1600–1667), English naval officer and Member of Parliament. The *OED3* gives the following definition for this usage of "datum": "Chiefly in *pl.* An item of (chiefly numerical) information,

A further extension to the "data" concept derives from the recognition that not all attribute-values are expressible in numeric form. Contemporary developments in semantic web technologies have demonstrated the utility in treating declarative statements as object–attribute–value (or subject–predicate–object) triples, whose values may take the form of character strings (or entity-names) of any kind; but the semantic web is just one recent manifestation of a broader understanding of "data" that became widespread early in the twentieth century. In the literature of the social sciences, the insight that data may be typed is sometimes attributed to the American psychologist Stanley Stevens (1906–73), citing his 1946 paper (or its 1941 presentation) as a source for a taxonomy of scales of measurement that distinguishes among nominal, ordinal, interval, and ratio scales (see *Figure 1*; Stevens, 1946; see also Newman, 1974). In developing (what Joel Michell calls) a representationalist theory of measurement (see, for example, Michell, 1986), Stevens was indebted to earlier writers such as the German physicist Hermann von Helmholtz (1821–94) and the British philosopher Bertrand Russell (1872–1970; see, for example, Helmholtz, 1977 and Russell, 1903). In fact, the distinction between ordinal and cardinal numbers can be traced back to the sixteenth century, that between qualitative and quantitative analysis to the early nineteenth, and that among kinds of variable observed or measured to the late nineteenth. A qualitative variable is one whose possible values are not numeric, but still answer the question, "What kind of thing is this?" A nominal or categorical variable is one whose values are names of classes or categories.[19] An ordinal variable is one whose values may be ranked, but do not express the size of the differences among them. The values of cardinal variables indicate how many (predefined) units of measurement are counted.[20]

esp. one obtained by scientific work, a number of which are typically collected together for reference, analysis, or calculation."

19 For those who would characterize numerals, also, as names of classes, we might wish to add to our definition of a nominal variable the constraint that its values are not arrangeable in any meaningful, quantitative order. That nominal variables may always be *coded* numerically (e.g., blue = 0, red = 1) is neither here nor there.

20 The fine distinction that Stevens drew between interval and ratio scales is no longer widely applied.

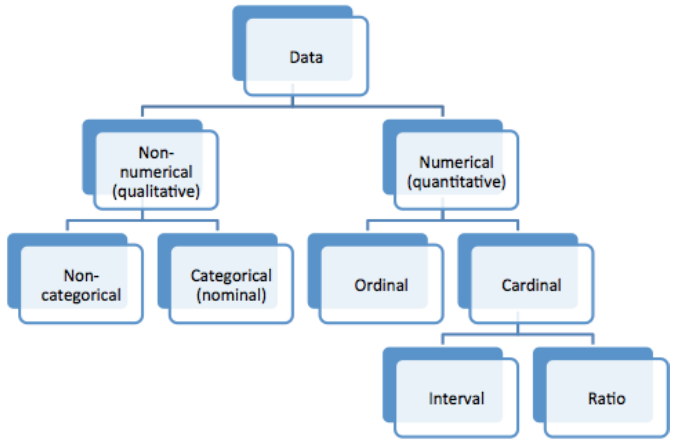

Fig. 1 A selection of data types

The essence of the representationalist view, according to Michell (1986, p. 398), is that "numbers are used in measurement to represent empirical relations between objects"; i.e., the numbers represent "an empirical relational system, which is thought of as an objective structure existing quite independent of our operations" (p. 404). The underlying idea that there exists an objective reality from which data may be "read off" is a cornerstone of the realist tradition in metaphysics. Positivist approaches to inquiry assume just such a realist ontology; conversely, constructivist approaches invoke a relativist ontology in which substantial components of the external world are understood to be constructed by, and dependent on the actions and subjective thoughts of, human agents.

Johanna Drucker (2011) argues that the very concept of "data" is meaningful only in the context of naïve realist approaches to inquiry. The realist's ontological assumption that phenomena exist independently of any observers—i.e., the assumption that any question of whether or not phenomenon X exists is a matter of objective fact—is typically accompanied by a set of premises as follows: that any question of whether or not phenomenon X has property Y is also a matter of objective fact; that, for any representation, description, or statement of the properties of phenomenon X to be considered true (correct, valid, etc.), that representation needs to correspond accurately to the facts; and that it is generally possible, under normal conditions and with appropriate measures taken to correct for known biases, to acquire knowledge of the facts through observation. In this context, the true statement, the act of unbiased observation, the facts, and the phenomenon itself are supposedly in such perfect alignment that they are treated almost as if they are all the same thing—with the effect of "collaps[ing] the critical distance between the phenomenal world and its interpretation, [and] undoing the basis of interpretation on which humanistic knowledge production is based." Specifically, data are conceived as no more than "mere descriptions" of a priori conditions.

Drucker notes that humanistic and interpretative inquiry acknowledges the "situated, partial, and constitutive" character of knowledge production that is ignored in the positivist's scientistic paradigm. The constructivist recognizes that "knowledge is … taken, not simply given as a natural representation of pre-existing fact." Drucker imagines the reconception of "data" as "capta" (the nominative neuter plural of the perfect passive participle of the third-conjugation Latin verb *capere*, "to take"): "while data are given, capta are taken."[21] For Drucker, this reconception is a necessary precondition of the "re-examin[ation of] the intellectual foundations of digital humanities" for which she calls.

Michell contrasts representationalism in measurement theory both with the operationalist view that measurement is simply "an operation that produces numbers" in some precisely specified way (p. 403)—from which we may infer that "the data on which measurement is based are *inherently* numerical" (p. 404, emphasis added)—and with the classical view, dating as far back as Aristotle and Euclid, that measurement is simply the assessment of quantity, i.e., the determination of "how much of a given attribute some object possesses" (p. 405).

1.8 The Computational Interpretation: Data as Bits (ca. 1980–)

The most recent stage in the evolution of standard senses of the "data" concept can be dated not to the 1950s or 1960s as one might expect, but to the last decades of the twentieth century. The earliest dictionaries of computing terminology did not mention "data" at all, and it was only with the production of the IBM 701 Electronic Data Processing Machine, launched in 1953, that the term began to be more widely used in a computing context. Initially, the sense was of data as attribute-values—the readings, measurements, and results of scientific inquiry, typically numeric in form, to be manipulated as "givens" for statistical and mathematical analysis—and this usage has of course persisted in the database community. In time, however, "data" has come to be regarded in some contexts as virtually synonymous with "bits" (i.e., binary digits—the 0s and 1s that computers handle or "process" at the most fundamental level), or "the digital," or even "the digitizable." A source of misunderstanding in contemporary discussions of data science and big data is a tendency to conflate three related but distinct interpretations: data as evidence; data as (typically numeric) attribute-values; and data as bits.

1.9 The Diaphoric Interpretation: Data as Differences (ca. 2000–)

A distinction similar to the one that Floridi draws among three types of diaphora (*de re*, *de signo*, and *de dicto*) may equally usefully be applied to an informational interpretation of "data." Just as classificatory analyses of "information" have pointed to the possibilities

21 Peter Checkland, developer of Soft Systems Methodology (SSM), seems to have been the first to suggest the term "capta," in the early 1980s (see, for example, Checkland, 1999, pp. 53–54); he uses it to denote "those items of data which we focus on, have a concern about, define, and select."

of finding informative entities in the worlds of physical reality, thought, and expression (see, for example, Furner, 2004), we might consider that data, too, may be found at the dedomenal (or noumenal), haireomenal (or phenomenal), and graphomenal levels:[22]

- data as objective reality, which we might determine (*contra* Kant, but in line with realist or representationalist accounts of measurement) is accessible to, and knowable and describable by, human agents;
- data as subjective appearances, observations, ideas, meanings, or propositional content; and
- data as linguistic expressions of individual observations.

For example, instead of (or in addition to) merely claiming that the sentence "Lubetzky is wise" is data, we might prefer to treat the mental image that we have of Lubetzky's wisdom as data, or even to treat the fact or state of Lubetzky's being wise as data. These distinctions replicate some that are frequently noted in discussions of the concept of information.

2 "Data" vs. "Document"

The most widely cited analysis of the concept of "document" is Michael Buckland's. Buckland first elucidates the notion of "information as evidence, as things from which one becomes informed," identifying three "sorts of things this might include": data; "text and documents"; and objects (Buckland, 1991, p. 353). He then suggests expanding the category of "document," along the lines taken by Otlet and Briet, to include not only texts but also all other physical things that are informative. For Buckland, the terms "document," "information as evidence," "information-as-thing," and "information resource" are equivalent; he argues that handwritten manuscripts, printed books, museum objects, tables of numerical data, and catalog records comprise separate "subsets" of the universal set of documents.[23] Using the terminology that Floridi applies to conceptions of "data," we may characterize Buckland's analysis as an instance of an informational interpretation of "document": for Buckland, the most salient feature of a document is not that it is a set of known facts from

22 In ancient Greek, the word *νοούμενον* (*nooúmenon*) is the neuter nominative singular of the passive present participle of the verb *νοέω* (*noéō*, "I know"); *αἱρεόμενον* (*haireómenon*) is the corresponding form of the verb *αἱρέω* (*hairéō*, "I take"); *φαινόμενον* (*phainómenon*) is derived from *φαίνω* (*phaínō,* "I appear"*)*; and *γραφόμενα* (*graphómena*) is derived from *γράφω* (*gráphō*, "I write"). In the philosophy of Immanuel Kant (1724–1804), a noumenon is a "thing-in-itself" (*Ding an sich*)—an object or event as it exists, independent of any sentient being—while a phenomenon is a thing as it is knowable by the senses.

23 The conception of information as evidence that Buckland presents here is different from that developed in the Schellenbergian tradition of archival science, where archival records serve as evidence of the contexts in which they were produced and used.

which as yet unknown facts may be deduced, nor that it is an aggregation of binary digits, but that it has propositional content.

One aspect of Buckland's original presentation of the documentalists' traditional understanding of "document" that is not quite clear is the precise nature of the relationship purported to exist between "information" and "data." Buckland notes that "data" is "an apt term for the sort of information-as-thing that has been processed in some way for use" (1991, p. 353), that it is also "commonly" used to denote "whatever records are stored in a computer" (1991, p. 353), and that there is "a tendency" to use it to denote "numerical information" (1991, p. 354). He concludes, however, that "it is wise not to assume any firm distinction between data, document, and text" (1991, p. 354).

The issue here is that it seems logically possible, at least for any alphanumeric text, to make conceptual distinctions among the following entities:[24]

1. any one of a number of material instances of the carrier of the text (each instance comprising an aggregation of, e.g., leaves of paper bound as a book);
2. the abstract form of the carrier of the text;
3. any one of a number of material instances of the symbolic encoding of the text (each instance comprising an aggregation of, e.g., characters, numerals, words, and/or values);
4. the abstract form of the encoding of the text;
5. any one of a number of material instances of the content of the text (each instance comprising an aggregation of, e.g., propositions); or
6. the abstract form of the content of the text.

Which (or which combination) of these entities, if any, is correctly labelled "document"—and, therefore, "information"—according to Buckland's model? The option that seems to fit the bill most closely is #5, interpreted as a physical thing that is informative. But then where, if anywhere, is "data" to be found among these entities?

The situation is complicated slightly by the difference in grammatical status of "document," "information," and "data."[25] Partly as a result, ordinary contemporary practice is to treat documents as discrete aggregations of quantities or amounts (rather than pieces or items) of information. Nevertheless, in general, the documentalist's understanding of the relationship among these entity-types may be modeled as follows (model A), combining an informational interpretation of "document" (as an aggregation of triples) with an

24 Cf. the items, manifestations, expressions, and works defined in the *Final report* of the IFLA Study Group on the Functional Requirements for Bibliographic Records (IFLA 1998): entities #1, #3, and #5 may be equated with items, entity #2 with a manifestation, #4 an expression, and #6 a work.

25 The first is a count noun; the second is a mass noun; and the third is sometimes used as a count noun, sometimes as a mass noun. It makes sense to talk of "ten documents," just as it does to talk of "ten dollars"; but we would not usually choose to talk of "ten informations," just as we would not normally say "ten monies." "Data," of course, is an odd case: its origin in Latin as a plural form leads many writers to insist on its taking a plural verb (e.g., "The data are …"), but presumably even those scholars would balk at using formulations like "ten data."

informational interpretation of "data" (as numeric attribute-values), with the upshot that "dataset" is treated as a species of "document":[26]

Documents consist of **Information**.
.. **Texts** (i.e., Textual documents) consist of **Text** (i.e., Textual information).
.. **Datasets** (i.e., Numerical documents) consist of **Data** (i.e., Numerical information).

If the inclusion of information as a separate entity in this model, with each of text and data considered as its species, seems to flout Occam's razor (see, for example, Furner, 2004), we might consider reformulating it as follows (B):

Document-sets consist of **Documents**.
.. **Texts** (i.e., Textual document-sets) consist of **Textual documents**.
.. **Datasets** (i.e., Numerical document-sets) consist of **Numerical documents**.

Each of these models stands in contrast to an understanding that is sometimes articulated in the computer science community, which combines a computational interpretation of "data" (as bits) with a computational interpretation of "document" (as bit-string) to arrive at a model (C) in which a document is treated as a species of dataset:

Datasets consist of **Data**.
.. **Documents** (i.e., Textual datasets) consist of **Textual data**.
.. **Numerical datasets** consist of **Numerical data**.

A variant (D) of this latter position rests on an intuition that, conceptually, there is no distinction to be made between "data" and "dataset"—that any given dataset is itself a set of datasets, and that any given datum is just as much an aggregation of data as any given dataset (and has just the same kinds of qualities):

Datasets consist of **Datasets**.
.. **Documents** (i.e., Textual datasets) consist of **Documents** (i.e., Textual datasets).
.. **Numerical datasets** consist of **Numerical datasets**.

Applying the same kind of thinking to the documentalist's position—recognizing that there is no distinction to be made between "document" and "document-set"—gives the following result (E):

26 In the presentation of this and succeeding models, the double-dot notation indicates a genus–species relationship. In model A, for example, each of the two entity-types **Texts** and **Datasets** is a species of **Documents**: in other words, all texts are documents, and all datasets are documents, but not all documents are texts, neither are all documents datasets.

Documents consist of **Documents**.
.. **Texts** (i.e., Textual documents) consist of **Texts** (i.e., Textual documents).
.. **Datasets** (i.e., Numerical documents) consist of **Datasets** (i.e., Numerical documents).

One of Allen Renear's long-term projects has been to refine a theory of documents that satisfactorily accounts for problems with characterizations of documents as aggregations, strings, graphs, relations, ordered hierarchies, tuples, sentences, or (in general) sets (see, for example, Coombs, Renear, & DeRose, 1987; DeRose, Durand, Mylonas, & Renear, 1990; Renear, Mylonas, & Durand, 1996; Renear, 2004; Renear & Wickett, 2010). With Simone Sacchi, Karen Wickett, and David Dubin, Renear has also developed a conceptual model in which the term "datasets" is used for "symbol structures that express data content together with, in many cases, auxiliary information" (Renear, Sacchi, & Wickett, 2010; Sacchi, Wickett, Renear, & Dubin, 2011; Wickett, Sacchi, Dubin, & Renear, 2012).

Renear and his colleagues distinguish carefully between, on the one hand, the data and metadata that are aggregated to form the "symbol structures" known as datasets, and on the other, the "content" or information that is expressed by those data and metadata, so that if we were to represent Renear's model in the form used in the previous section, part of it would look like this (F):

Datasets consist of **Data** and **Metadata**.
.. **Data** express **Data content**.
.. **Metadata** express **Contextual information about data**.

One supposed benefit of this approach is that we can represent, in the model itself, the classic semiotic distinction between signifier and signified, sentence and proposition, term and concept, expression and thought, signal and message, and (perhaps) data and information, where the information "contained in" or communicated by a dataset is equivalent to an aggregation of the meanings attributed to that dataset by the consumer. Simplifying our presentation of Renear's model in one respect (flattening the distinction between data and metadata) and augmenting it in another (reintroducing the distinction between the textual and the numerical), we might arrive somewhere like this (G):[27]

Documents consist of **Data**, which express **Information**.
.. **Texts** (i.e., Textual documents) consist of **Text** (i.e., Textual data).
.. **Datasets** (i.e., Numerical documents) consist of **Numerical data**.

27 One feature of Renear's model is its generality, such that the category of **Datasets** is conceived as inclusive not only of textual and numerical artefacts, but also of pictorial, audiovisual, and multimedia forms (among potentially many others). So to reintroduce the textual–numerical distinction at this point might be construed, not so much as a desirably particularizing move, as an undesirably limiting one. But the identification of these two species is intended to be representative rather than exhaustive of all possibilities. I am grateful to Johanna Drucker for pointing out the need for clarification here.

In a final iteration, we can take on board the documentalist's idea of "document," rather than "data," as the broadest category of triple-aggregations (as applied in model A), and the notion that there is no useful conceptual distinction to be made between "document" and "document-set" (as applied in model E), with the following result (H):

Documents consist of **Documents**, which express **Information**.
.. **Texts** (i.e., Textual documents) consist of **Texts** (i.e., Textual documents).
.. Datasets (i.e., Numerical documents) consist of **Datasets** (i.e., Numerical documents).

This, then, is the result of combining an informational interpretation of "data," an informational interpretation of "document," and a subjectivist interpretation of "information" (as meaning). It is formally very similar to the original documentalist's position presented above, and draws a real and useful distinction between "document" (interpreted as denoting entities of type #3 in the list above) and "information" (reserved in this context for entities of type #5).

3 Conclusion

The downsides to the computational interpretation of "data" and "document" (see Section 2.8) are several.

- It is incorrect. It is not the case that all the entities that play the roles to which we assign the names "data" and "document" are digital or even digitizable.
- It is incoherent. The binary digits that are said to comprise documents are of a kind that is indistinguishable from the kind of binary digits that are said to comprise data. So how can the distinction between documentary data and non-documentary data be made?
- It is useless.[28] There is nothing conceptually to be gained by reducing the entire universe of cultural, scientific, creative, and intellectual works to a sequence of 0s and 1s.

The informational interpretation, on the other hand (see Section 2.7), comes warmly recommended. It allows us to give distinctive names to a number of simple concepts to which we need to refer frequently and extensively in both day-to-day and specialized activities, and it has had reputable adherents for centuries. It is not in fact the case that documents are made up of data, nor that the document is a species of dataset: rather it is the other way round, in both respects. A dataset is made up of documents; and the dataset is a species of document.

28 This is the most significant downside. If the computational interpretation were useful, it wouldn't matter so much that it's incorrect.

References

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*, 3–9.

Bailey, N. (1721). *An universal etymological English dictionary*. London, England: Printed for E. Bell et al.

Balbus, J. (1971). *Catholicon*. Farnborough, England: Gregg. [Facsimile reprint of the 1st printed ed. (1460) of the thirteenth-century text, printed in Mainz. See http://lcweb2.loc.gov/service/rbc/rbc0001/2012/2012rosen0031/2012rosen0031.pdf for a digitized copy.]

*Batten, Sir W. (1630). *A most plaine and easie way for the finding of the sunnes amplitude and azimuth, and thereby the variation of the compasse, by logarithme*. N.p.: E. Allde for J. Tapp.

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge, MA: MIT Press.

Borgman, C. L. (2014). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

*Boughen, E. (1650). *Master Geree's Case of Conscience Sifted*. London, England.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, *42*(5), 351–360.

Buckland, M. K. (1997). What is a "document"? *Journal of the American Society for Information Science*, *48*(9), 804–809.

Capurro, R., & Hjørland, B. (2003). The concept of information. *Annual Review of Information Science and Technology*, *37*, 343–411.

†‡Chambers, E. (1728). *Cyclopædia: or, An universal dictionary of arts and sciences*. London, England: Printed for J. & J. Knapton et al.

Checkland, P. (1999). Systems thinking. In W. Currie & B. Galliers (Eds.), *Rethinking management information systems: An interdisciplinary perspective* (pp. 45–56). Oxford, England: Oxford University Press.

Coombs, J. H., Renear, A. H., & DeRose, S. J. (1987). Markup systems and the future of scholarly text processing. *Communications of the ACM*, *30*(11), 933–947.

Day, R. E. (2001). *The modern invention of information: Discourse, history, and power*. Carbondale: Southern Illinois University Press.

DeRose, S. J., Durand, D. G., Mylonas, E., & Renear, A.H. (1990). What is text, really? *Journal of Computing in Higher Education*, *1*(2), 3–26.

*Donne, J. (1649). *Fifty sermons*. London, England: Printed by J. Flesher for M. F. et al.

Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, *5*(1). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Eliot, T. S. (1934). *The rock*. New York, NY: Harcourt, Brace.

¶*Encyclopædia Britannica: or, A dictionary of arts and sciences, compiled upon a new plan*. (1771). Edinburgh: Printed for A. Bell and C. Macfarquhar.

*Euclid. (1661). *Euclid's* Elements *of geometry: … Likewise Euclid's* Data. J. Leeke & G. Serle (Eds.). London, England: Printed by R. & W. Leybourn for R. Tomlins.

Ferraris, M. (2013). *Documentality: Why it is necessary to leave traces*. Bronx, NY: Fordham University Press.

Floridi, L. (2008). Data. In W. A. Darity, Jr. (Ed.), *International encyclopedia of the social sciences* (2nd ed., vol. 2, pp. 234–237). Detroit, MI: Macmillan.

Frické, M. (2009). The knowledge pyramid: A critique of the DIKW hierarchy. *Journal of Information Science*, *35*(2), 131–142.

Frohmann, B. (2004). *Deflating information: From science studies to documentation*. Toronto, Canada: University of Toronto Press.

Furner, J. (2004). Information studies without information. *Library Trends*, *52*(3), 427–446.

Furner, J. (2014). Information without information studies. In F. Ibekwe-SanJuan and T. M. Dousa (Eds.), *Theories of information, communication and knowledge* (pp. 143–179). Dordrecht, The Netherlands: Springer Science+Business Media.

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

Gitelman, L. (2014). *Paper knowledge: Toward a media history of documents*. Durham, NC: Duke University Press.

Glare, P. G. W. (Ed.). (2012). *Oxford Latin dictionary* (2nd ed.). Oxford, England: Oxford University Press.

*Hammond, H. (1647). *A copy of some papers past at Oxford, betwixt the author of the* Practicall catechisme*, and Mr. Ch*. London, England: Printed by R. Cotes for R. Royston.

*Hammond, H. (1648). *A brief vindication of three passages in the* Practical catechisme. London, England: Printed for R. Royston.

‡Harris, J. (1704). *Lexicon technicum: or, An universal English dictionary of arts and sciences*. London, England: Printed for D. Brown et al.

Helmholtz, H. von. (1977). *Epistemological writings: The Paul Hertz / Moritz Schlick centenary edition of 1921*. R. S. Cohen & Y. Elkana (Eds.). Dordrecht, The Netherlands: D. Reidel.

Hortus vocabulorum, 1500. (1968). R. C. Alston (Ed.). Menston, England: Scolar. [Facsimile reprint of the 1st ed. (1500) printed by W. de Worde, London.]

∫Huntsman, J. F. (1973). "Pepys MS 2002' *Medulla grammatice*: An edition." PhD diss. ProQuest (7326020).

IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). *Functional requirements for bibliographic records: Final report*. Munich, Germany: K. G. Saur.

†‡Johnson, S. (1755). *A dictionary of the English language*. London, England: Printed by W. Strahan, for J. & P. Knapton et al.

Kent, W. (1978). *Data and reality: Basic assumptions in data processing reconsidered*. Amsterdam, The Netherlands: North-Holland.

Latham, R. E. (1986). *Dictionary of medieval Latin from British sources* (vol. 1: A–L, fasc. 3). London, England: British Academy / Oxford University Press.

Lund, N. W. (2009). Document theory. *Annual Review of Information Science and Technology*, 43, 1–55.

Markham, A. N. (2013). Undermining "data": A critical examination of a core term in scientific inquiry. *First Monday*, *18*(10). http://firstmonday.org/article/view/4868/3749

†Martin, B. (1749). *Lingua Britannica reformata: or, A new English dictionary*. London, England: Printed for J. Hodges et al.

Michell, J. (1986). Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, *100*(3), 398–407.

*Moxon, J. (1679). *Mathematicks made easie: or, A mathematical dictionary*. London, England: Printed for J. Moxon.

Newman, E. B. (1974). On the origin of "scales of measurement." In H. R. Moskowitz, B. Scharf, & J. C. Stevens (Eds.), *Sensation and measurement* (pp. 137–145). Dordrecht, The Netherlands: D. Reidel.

†Ogilvie, J. (1882). *The imperial dictionary of the English language: A complete encyclopedic lexicon, literary, scientific, and technological* (new ed.). C. Annandale (Ed.). London, England: Blackie.

Renear, A. H. (2004). Text encoding. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A companion to digital humanities* (pp. 218–239). Oxford: Blackwell.

Renear, A. H., Sacchi, S., & Wickett, K. M. (2010). Definitions of *dataset* in the scientific and technical literature. *Proceedings of the American Society for Information Science and Technology*, 47. http://onlinelibrary.wiley.com/doi/10.1002/meet.14504701240/full

Renear, A. H., Mylonas, E., & Durand, D. G. (1996). Refining our notion of what text really is: The problem of overlapping hierarchies. In S. Hockey & N. Ide (Eds.), *Research in humanities computing 4: Selected papers from the ALLC/ACH Conference, Christ Church, Oxford, April 1992* (pp. 263–280). Oxford, England: Clarendon Press.

Renear, A. H., & Wickett, K. M. (2010). There are no documents. In *Proceedings of Balisage: The Markup Conference 2010*. http://www.balisage.net/Proceedings/vol5/print/Renear01/BalisageVol5-Renear01.html

Riles, A. (Ed.). (2006). *Documents: Artifacts of modern knowledge*. Ann Arbor, MI: University of Michigan Press.

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

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180.

Russell, B. (1903). *The principles of mathematics*. Cambridge, England: Cambridge University Press.

Sacchi, S., Wickett, K. M., Renear, A. H., & Dubin, D. (2011). A framework for applying the concept of significant properties to datasets. *Proceedings of the American Society for Information Science and Technology*, *48*. http://onlinelibrary.wiley.com/doi/10.1002/meet.2011.14504801148/full

Simpson, J. (Ed.). (2000–). *Oxford English dictionary (3rd ed.). Oxford, England: Oxford University Press. http://www.oed.com/*

Simsion, G. C. (2007). *Data modeling: Theory and practice*. Denville, NJ: Technics Publications.

Smith, B. (2014). Document acts. In A. Konzelmann Ziv & H. B. Schmid (Eds.), *Institutions, emotions, and group agents: Contributions to social ontology* (pp. 19–31). Dordrecht, The Netherlands: Springer Science+Business Media.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*(2684), 677–680.

Thomas, T. (1972). *Dictionarium linguae Latinae et Anglicanae, 1587*. R. C. Alston (Ed.). Menston, England: Scolar. [Facsimile reprint of the 1st ed. (1587) published by the author in Cambridge.]

Thow-Yick, L. (1994). The basic entity model: A fundamental theoretical model of information and information processing. *Information Processing & Management*, *30*(5), 647–661.

∫Tremblay, F. A. (1968). "The Latin–Middle English glossary 'Medulla grammatice,' B. M. Harl. MS 1738." PhD diss., Catholic University of America. ProQuest (6912308).

*Tuke, T. (1614). *Nevv essayes: meditations, and vowes*. London, England: Printed by N. Okes.

*Urquhart, T. (1645). *The trissotetras: or, A most exquisite table for resolving all manner of triangles*. London, England: Printed by I. Young.

∫Van Zandt-McCleary, J. M. (1958). "The *Medulla grammatice* Latin–English dictionary: A diplomatic transcription." PhD diss., University of Chicago. ProQuest (T-04429).

†Webster, N. (1828). *An American dictionary of the English language*. New York, NY: S. Converse.

Wickett, K. M., Sacchi, S., Dubin, D., & Renear, A. H. (2012). Identifying content and levels of representation in scientific data. *Proceedings of the American Society for Information Science and Technology*, *49*. http://onlinelibrary.wiley.com/doi/10.1002/meet.14504901199/full

†Worcester, J. E. (1860). *A dictionary of the English language*. Boston, MA: Swan, Brewer & Tileston.

Zeleny, M. (1987). Management support systems: Towards integrated knowledge management. *Human Systems Management*, *7*(1), 59–70.

Electronic versions of some of the works listed may be found according to the following key:

**Early English Books Online*.
http://eebo.chadwyck.com/home/ or http://quod.lib.umich.edu/e/eebogroup/.
¶*Eighteenth Century Collections Online*.
http://find.galegroup.com/ecco/ or http://quod.lib.umich.edu/e/ecco/.
†*HathiTrust Digital Library*.
http://www.hathitrust.org/.
‡*The Making of the Modern World*.
http://find.galegroup.com/mome/.
∫*ProQuest Dissertations & Theses Global*.
http://search.proquest.com/pqdtglobal.