

Scientists' Reuse of Old Empirical Data: Epistemological Aspects

James W. McAllister*†

This article investigates epistemological aspects of scientists' reuse of empirical data over decades and centuries. Giving examples, I discuss three respects in which empirical data are historical entities and the implications for the notion of data reuse. First, any data reuse necessitates metadata, which specify the data's circumstances of origin. Second, interpretation of historical data often requires the tools of humanities disciplines, which produce a further historicization of data. Finally, some qualitative social scientists hold that data are personal to the researcher who coconstructs them in the research process and are therefore skeptical about the prospects of reusing data.

1. Introduction. This article is about scientists' reuse of empirical data—the results of observations and measurements. An event takes place at time t_1 ; researchers capture (and, perhaps, first use) empirical data about that event at t_2 ; the same or (more typically) other researchers then reuse those data at t_3 . Times t_1 and t_2 may coincide, but t_3 is later than t_2 . The aim of the reuse at t_3 may be to answer with improved analytical techniques the research question for which investigators gathered the data at t_2 or to answer new questions.

There are many studies of the swift reuse of empirical data. This article, by contrast, focuses on cases in which the interval between t_2 and t_3 is long—decades or centuries. In these cases, the historical and intellectual contexts in which the data are used differ from those in which they were collected. The historically remote origin of the data when they are reused gives rise to epistemological issues, as we shall see.

*To contact the author, please write to: Institute of Philosophy, University of Leiden, PO Box 9515, 2300 RA Leiden, The Netherlands; e-mail: j.w.mcallister@phil.leidenuniv.nl.

†My thanks to Sabina Leonelli for organizing the symposium "Data in Time: The Epistemology of Historical Data," 25th Biennial Meeting, Philosophy of Science Association, Atlanta, November 2016; the other speakers, Rachel Ankeny, David Sepkoski, and Alison Wylie; the audience for stimulating discussion; and Wendy Parker and two unnamed referees for comments on a previous draft.

Philosophy of Science, 85 (December 2018) pp. 755–766. 0031-8248/2018/8505-0002\$10.00
Copyright 2018 by the Philosophy of Science Association. All rights reserved.

Reuse of old data seems to serve several admirable epistemic values and to embody an optimistic view of science. It harnesses previous scientific labor, maximizing the value of existing data; it extends the scope of human cognitive access to phenomena; it embodies positivist ideals of continuity of scientific practice and cumulativeness of empirical findings; it seems to assure objectivity, by insulating the process of data gathering from the purposes of a subsequent research project; and it even carries a suggestion of tapping into age-old insights and wisdom. For these reasons, many iconic success stories in science feature reuse of old data.

Old empirical data are especially important in research in historical sciences into unique or infrequent events and long-term trends. Astronomy has drawn on historical data to study rare events and secular changes in celestial objects (Stanley 2013; Hsia 2017). Halley's comet is a paradigmatic example. Edmund Halley's use at the beginning of the eighteenth century of the observations of Johannes Kepler in 1607, Peter Apian in 1531, and others to predict the comet's return in 1758 provided one of the triumphs of early Newtonianism (Cook 1998, 214). Subsequently, the recovery of observational records going back to ancient Chinese and Babylonian times has enabled astronomers to refine estimates of the comet's orbital parameters (Kronk 1999, 5–6, 8–10).

At least two current developments have rendered understanding the epistemology of the reuse of decades- and centuries-old data more urgent. The first is efforts to reconstruct climate history. Our knowledge of past climate is based mainly on present-day observations of physical traces, such as ocean floor sediments and ice cores, but climate scientists gladly reuse past direct observations where they can. The Central England Temperature record, stretching back to 1659, is the best-known example, but researchers have recovered further archives from time to time (Jones 2008; Sharma et al. 2016). The second development is current projects to preserve historical data sets for future researchers. One of the largest is Digital Access to a Sky Century at Harvard (DASCH), which aims to digitize 500,000 photographs of the sky in the form of glass plates exposed between 1885 and 1992 (Grindlay et al. 2009). There have been many calls for further such initiatives to preserve both quantitative and qualitative empirical data (e.g., Griffin and the CODATA Task Group 2015; Bishop and Kuula-Luumi 2017; Griffin 2017). Both these developments heighten the need to understand what is involved in reusing legacy data.

This article discusses difficulties and limitations of the reuse of decades- and centuries-old empirical data. I focus on methodological problems that arise from the historicity of empirical data, or the respects in which data are historical constructs. Empirical data are historical entities, both because the significance of data depends on the circumstances of their origin and because data take a position not just epistemically but also historically intermediate between phenomena in the world and investigators. As Leonelli (2018) has

emphasized, scientists' understanding of the temporality of data affects their interpretations of and inferences from those data.

Mainstream philosophy of science has been reluctant to view empirical data as historical. Even in its postpositivist period, the discipline has tended to conceptualize context dependence of data mainly in terms of theory ladenness, or the extent to which observation depended on an investigator's theoretical presuppositions. For example, Wylie (2017, 221n6) has credited "Hanson's classic account of the theory-ladenness of observation" as part of "the intellectual background" of the discussion of limitations on the reuse of old data in archaeology.

In fact, theory ladenness is not a big conceptual challenge to reusing empirical data. Whereas theory ladenness suggests that researchers with new theoretical presuppositions will interpret data differently, it leaves data fully available for reuse under such new presuppositions—as Hanson's own bird-antelope metaphor suggests (1958, 13–15). Accordingly, there are many historical cases in which researchers have used old data under new presuppositions. One intriguing way is by exploiting components of data sets that the original researchers had dismissed as noise. In 1964, for example, Robert H. Dicke, Jim Peebles, and David Wilkinson reused Arno Penzias and Robert Wilson's microwave data, collected to calibrate a communications antenna, to confirm their hypothesis about cosmic background radiation (Wilkinson and Peebles 2000).

In section 2, I discuss the importance of metadata. I consider the role of humanities disciplines in the reuse of centuries-old empirical data in section 3. In section 4, I discuss a recent debate about secondary analysis of qualitative data in methodology of social science. Powerful views advanced in that debate cast doubt not just on the feasibility of data reuse but also on the very coherence of the concept.

2. Metadata and Evidential Significance. When a research team gathers and then uses a body of data within a short time span, it may expect to retain firsthand knowledge about its origin. When collection and reuse of data are separated by decades or centuries, however, the subsequent researchers must rely on metadata.

Metadata consist of information about the origin and handling of an empirical data set (Edwards et al. 2011; Vanderbilt and Blankman 2017). The content of a data set—a certain string of digits, for example—has no determinate evidential significance separate from an assurance that it was gathered in specific circumstances, for example, by a certain researcher using a certain instrument at a certain place and time to sample a certain physical phenomenon. Only the circumstances of origin distinguish a genuine empirical data set from a plausible fabrication, to mention the most extreme case. The metadata accompanying a data set provide this background information.

The strategy of relying on metadata is vulnerable to two problems, however. First, metadata are liable to loss, partly because, in many cases, they look like ephemeral annotations with a status lower than data. Many current initiatives for recovery of empirical data, including the DASCH project, explicitly devote much effort to preserving metadata too. The metadata accompanying the photographic plates digitized by DASCH take the form of handwritten notes in logbooks and on paper sleeves in which the plates were stored. They record the telescope used and its location, pointing direction, time and exposure of the observation, name of the object being observed, name of the observer, and the like (Schechner and Sliski 2016).

Second, subsequent reusers of empirical data may need items of metadata that the original collectors did not think of registering. This is a problem particularly if subsequent investigators reuse data to tackle a new research question, which requires different background information.

The following case of data reuse illustrates both problems. R. Elizabeth Griffin reanalyzed historical data from ground-based stellar spectra observations to answer a research question in a different branch of science: how ozone levels in the atmosphere varied during the twentieth century. Ground-based stellar spectrographs show absorption lines due to atmospheric ozone. The new study recovered ozone signatures from 16 photographic spectrograms—an obsolete storage format—made in the 1930s and 1940s and preserved at Mount Wilson Observatory, California.

For one thing, metadata were not always available, as Griffin reported: “loss (or temporary disappearance) of the early Mount Wilson log-books unfortunately caused difficulties when searching for appropriate spectra” (2006, 2232). In addition, however, the new project required particular items of metadata. The magnitude of the atmospheric effects, unlike the stellar spectra, depended on the path length through the atmosphere and thus on the altitude of the star. If an exposure was interrupted by clouds, for example, and resumed when the star was at a different altitude, the effective path length would vary. Correctly interpreting a spectrogram in the new project thus depended on the metadata recording the altitude of the star—a circumstance that the original project in stellar spectroscopy might well have regarded as irrelevant. Despite all difficulties, partly by calibrating the measurements against contemporaneous satellite data, Griffin validated the procedure, concluding that it could be applied with reasonable confidence also to even older stellar spectra data.

Metadata are important also when scientists reanalyze physical specimens collected long ago. The reanalysis of old specimens yields new empirical data, which originate at the time of the reanalysis. In order to understand the significance of both old specimens and the data gathered in the reanalysis, however, we need contextual information about the circumstances in which the specimens were selected, acquired, handled, and preserved. Historical phenology provides an example (Willis et al. 2017). Scientists can use herbarium

specimens to determine past flowering times and thereby reconstruct the response of plant species to climate change, but only if they have accurate contemporaneous data on where and when the specimens were collected. Robbirt et al. (2011) examined the field notes accompanying 192 orchid specimens collected between 1848 and 1958 preserved at the Natural History Museum, London, and Royal Botanic Gardens, Kew. They found inadequate records for almost half the specimens, such as missing, imprecise, unclear, or illegible records of collection date. Nonetheless, they judged that the use of herbarium collections to ascertain the effect of climate on plant phenology was validated.

As the examples in this section have shown, a first form of historicity of empirical data, which is captured by the concept of metadata, consists in the fact that their meaning and significance depend on the circumstances of their origin. But there is more.

3. Humanities Disciplines in Data Reuse. Much centuries-old evidence takes the form of text or images created within practices different from modern science and in very different cultural settings. The interpretation of such evidence requires contextual knowledge and hermeneutic skills that are part of the tool box of humanities scholars. Many natural scientists have tended to gloss over methodological and interpretive problems involved in extracting observational records from these sources. In this section, I review some challenges of interpretation that humanities scholars would regard as substantial.

Astronomy provides many examples. First, Hisashi Hayakawa and colleagues investigated the hypothesis that a large solar flare occurred in the late tenth century. Evidence for this hypothesis included sharp increases of carbon-14 in tree rings dating from 993 and 994, suggesting an increased cosmic ray flux on Earth. The research team searched “contemporary historical documents all over the world” (Hayakawa et al. 2017, 3) for references to red auroras, which would follow a geomagnetic storm. They found such mentions in five records from Saxony, two from Ireland, and one from the Korean peninsula written in 992 and 993. For example, they quoted from a manuscript of an anonymous work, *Annales Quedlinburgenses*, preserved in the Saxon State and University Library Dresden: “DCCCCXCII . . . XII Calend: Novembris totum cœlom ter in nocte visum est rubrum fuisse,” which the research team translated as “992 . . . On 10.21, the whole sky was seen reddened three times” (4). The research team interpreted such statements as straightforward eyewitness testimony but devoted little attention to the difficulties of interpreting early medieval manuscripts from diverse cultural settings.

A second astronomical example pertains to historical novae, a topic on the border between astronomy and history of science (Stephenson and Green 2002, 2005). Michael M. Shara and colleagues used a March 11, 1437, entry in *Sejong Sillok*, annals of the reign of King Sejong of Korea, to trace a nova

in the constellation Scorpius: “19th year of King Sejong, 2nd lunar month, day *yichou* [the 2nd day of the 60-day cycle], A meteor (*liuxing*) appeared . . . A solar halo . . . A guest star (*kexing*) began to be (*shi*) seen between the second and third stars of Wei. It was nearer to the third star, about half a *chi* (‘half a foot’) away. It lasted (*jiu*) for 14 days” (Shara et al. 2017, “Methods”).

By interpreting the star counting convention, the unit of distance, and other elements of this entry, Shara et al. calculated the likely coordinates of the nova in the sky. They then identified an object at corresponding coordinates in present-day telescope images as the nova’s remnants. In a further instance of data reuse, they found the same object in a 1923 photographic plate that the DASCH project had digitized. Once more, scholars in humanities disciplines would regard the interpretation of a fifteenth-century text as requiring care: “When it comes to analyzing ancient records, it can be a challenge interpreting them correctly,” Shara acknowledged (quoted in Choi [2017]). The article, however, provided little detail of the team’s approach, not even giving the source of the English translation of *Sejong Sillok* that they used.

Volcanologists have often used old data in the form of contemporaneous lay reports and pictures to reconstruct historical volcanic events (Pyle 2017). Historical sources have played an especially important role in the effort to reconstruct ground rise and fall at the Campi Flegrei caldera near Naples (Guidoboni and Ciuccarelli 2011). Scholars since the early nineteenth century have noted a band of perforations by marine mollusks in the limestone columns of the Temple of Serapis, a Roman market building in the Campi Flegrei, concluding that they must have been submerged in the sea to a depth of at least 6 meters in one or more periods (Rudwick 2008, 106–13, 272–75, 297–300). Radiocarbon dating has provided estimates of the ages of the mollusk colonies, but with substantial uncertainties.

Geologists have therefore turned to historical records to refine the dating of the ground movements. For example, Bellucci et al. (2006, 149) used an illustration in a 1430 manuscript of Pietro da Eboli’s didactic poem, *De balneis Puteolanis* (The Baths of Pozzuoli), which depicted two classical columns protruding from the sea behind bathers in the thermal spring. Having identified these as belonging to the Temple of Serapis, the research team inferred that the temple floor lay 10 meters below sea level around 1430.

Bellucci et al. seem to have interpreted the illustration as a straightforward naturalistic depiction. Humanities scholars might well advise caution, however. Classical ruins were a common theme in quattrocento painting: Andrea Mantegna is a famous example (Forero-Mendoza 2002, 91–105). Such depictions were motivated often by something other than naturalism, such as an intention to symbolize the superiority of Christianity over paganism. Partly submerged Roman columns in an early-fifteenth-century illustration of a didactic poem might thus not be reliable evidence of the situation of such columns.

The possibility of misinterpretation of similar sources is more than theoretical. Emanuela Guidoboni reported a belief widespread in the volcanology literature that Solfatara, a shallow crater in the Campi Flegrei, erupted in 1198. Among the evidence for this belief was an illustration in another manuscript of *De balneis Puteolanis*, dating from the late thirteenth century, which seemed to show flames emanating from a mountaintop. Guidoboni pointed out that contemporaneous chronicles did not mention an eruption in 1198 and suggested that volcanologists might have misinterpreted the picture in question: it might show merely heat rising from a thermal spring rather than volcanic activity. As Guidoboni wrote, “naturalistic realism was rare if not nonexistent in medieval illustrations, as any historical interpretation must emphasize” (2010, 231). In this light, Guidoboni called for more intensive collaboration between volcanologists and historians in interpreting old sources.

Archaeology has a well-established practice of integrating material findings and historical written sources, but the treatment of literary evidence requires care. Charlotte Hedenstierna-Jonson and colleagues reexamined a body excavated in the 1870s from a Viking-age grave at Birka, Sweden. They claimed, on the basis of a genomic analysis of the skeleton, that it was of a high-ranking woman Viking warrior. The team drew in part on literary sources to support this striking claim: their report opened with references to early medieval poetry and artworks about female Viking fighters and closed with a stanza of the *Poetic Edda*, composed probably in Greenland in the twelfth century, about a “high-born lady” who “took a naked sword and fought for her kinsmen’s lives” (Hedenstierna-Jonson et al. 2017, 858).

A humanities scholar might inquire to what extent we can regard such literary sources as factual reports rather than as contributions to mythology. The authors, however, seemed to suggest that centuries-old texts spoke for themselves and required no critical interpretation or specialist expertise: the research team included no scholars of language or literature, as Jesch (2017) noted. Incidentally, this case also underlined the importance of metadata: the nineteenth-century archaeologists who excavated the grave did not properly label the findings, leaving some uncertainty as to their original location.

The research projects discussed in this section up to now originated in astronomy, volcanology, or archaeology and subsequently entered terrain typical of humanities disciplines. In some other cases, a project has arisen in humanities disciplines and migrated to natural sciences. For example, Emmanuel Le Roy Ladurie, a medievalist, founded a long and productive research program into historical climate with an article in the principal journal of the *Annales* school of history, a paradigmatic humanities venue (1959). As sources, Le Roy Ladurie used historical records of temperature and rainfall, the dates, quantities, and prices of grape, wheat, and other harvests, wine vintage quality, the movements of Alpine glaciers, and the like. In time, the natural science community took up this research program. For example,

Isabelle Chuine, an ecologist, led a team including Le Roy Ladurie that used records of grape harvest dates in Burgundy to reconstruct spring and summer temperatures since 1370, publishing their findings in *Nature* as a contribution to historical phenology (Chuine et al. 2004). Such a trajectory helps to ensure that research projects in historical natural sciences incorporate sufficient expertise of humanities disciplines from the outset.

4. Methodological Debates in Qualitative Social Science. The most radical and searching challenge to the cogency of the concept of data reuse is the view that empirical data are personal to the researcher who coconstructs them in the context of a research project and are unusable by any other researcher in any later project.

Qualitative social science, in broad terms, focuses on understanding how people think, feel, or behave in particular situations or in relation with others, the meanings that they attribute to different aspects of their lives, and how they understand their own and others' behavior and beliefs. Typical empirical data have a discursive form: transcripts of in-depth interviews, reports of participant observation and other ethnographic fieldwork, and informants' diaries and logs.

There are some celebrated examples of data reuse in qualitative social science—what social scientists call “secondary analysis.” For example, Paul Thompson initiated the first national oral history project in Britain, “Family Life and Work Experience before 1918,” in the 1970s. It comprised wide-ranging structured interviews with a nationwide sample of over 500 people born up to 1918. Thompson used the data for his own research project on the Edwardian family, but the archived transcripts were a rich enough source for other researchers to mine in some 20 further major publications on topics ranging from working-class culture to social mobility, which appeared over decades to follow.

Thompson (2000) acknowledged, however, that qualitative social scientists showed “reluctance to draw on material created by other researchers.” Some reasons may be practical, to do with the unstandardized format of qualitative data, and ethical, to do with confidentiality assurances made to respondents. Methodological concerns also play a role, however. Qualitative social scientists have engaged in a long debate about the concept of reuse of empirical data that is more sophisticated and critical than any discussion of the same topic in the natural sciences (for overviews, see Hammersley [2010] and Tarrant [2017]). The debate has addressed the difference between qualitative and quantitative approaches, often in the context of a critique of positivist frameworks, which many qualitative researchers think quantitative researchers endorse.

Many qualitative social scientists hold a concept of “empirical data” radically different from that common in the natural sciences or even in quanti-

tative social science. This is true particularly for approaches influenced by the interpretive turn in social science, which drew on insights from ethnomethodology, phenomenology, and hermeneutics (Rabinow and Sullivan 1987). In these approaches, qualitative data are not “given” observations of external social facts that are “out there” independent of the researcher and available to be collected. The researcher plays a central role not only in interpreting data but also in constructing them. Qualitative data are derived from and are dependent on the relationship between a researcher and respondents. These actors coconstruct qualitative data through interpersonal relations within a research project (Law 2004). Data collection is therefore no passive extraction of information from participants by the researcher, but rather a joint construction of meaning.

A good example is the in-depth interview, in which the researcher and the respondents together create the outcome. A different researcher, perhaps with different theoretical commitments or cultural background, would have led to different data. As Mauthner, Parry, and Backett-Milburn put it, “‘findings’ are not *in* the data but *created* through the interaction of particular . . . researchers with particular respondents in particular locations and at particular historical junctures” (1998, 735).

This has consequences, first, for a researcher’s use of own data. To be able to make sense of data that he or she has produced, for example, in the interview setting, the researcher must review his or her own experience of that process and critically reevaluate the role that he or she played in the production of the data.

Even greater consequences follow for the concept of “data reuse.” The constructed nature of qualitative data and its dependence on the context of their production make it difficult to give any content to the concept of reusing someone else’s data. The researcher attempting secondary analysis of data faces three limitations: he or she has no personal relationship with the respondents and so will find it difficult to understand the data; he or she did not co-create the data, so the data will not incorporate his or her contribution; and he or she was not present at the production of the data and so has no firsthand awareness of the role that the original researcher played in the cocreation. In sum, data removed from the context of their production and from the original researcher and used by others do not have evidential force in qualitative projects. As Hammersley has put it, data produced by different researchers “cannot be treated as if they represent a common currency” (1997, 139). Calls for greater reuse of qualitative data thus seem to belong to an outdated neopositivistic project, according to Slavnić (2013).

This helps us to understand controversies about data reuse in qualitative social science. Heaton introduced a monograph on the topic as follows: “The first and most rudimentary principle of secondary analysis is that it involves the use of *pre-existing* data” (2004, 2). There may be innocuous senses of

“pre-existing,” but Moore (2007) has objected to the term as suggesting that qualitative data were available at the outset of a research project rather than being coproduced in the research process. Moore has sketched an alternative view in which what we call “secondary analysis” starts with recontextualizing and reconstructing data. The entity thereby produced is not “secondary data” for which we need a method of “secondary analysis,” but rather a new order of data that merits its own primary analysis.

5. Conclusions. I have looked at three respects in which empirical data are historical entities and at repercussions of these respects for the notion of reuse of old data. First, since the evidential significance of data is not assured without information about the circumstances of the data’s origin, metadata are an essential accompaniment to any data reuse. Second, since much data reuse over centuries relies on text and pictures produced outside scientific discourse, researchers must contextualize those sources using tools characteristic of humanities disciplines. Finally, debates in methodology of qualitative social science attribute an extreme historical specificity to empirical data, ruling out any form of reuse outside the context of coproduction of the data.

REFERENCES

- Bellucci, Francesca, Judith Woo, Christopher R. J. Kilburn, and Giuseppe Rolandi. 2006. “Ground Deformation at Campi Flegrei, Italy: Implications for Hazard Assessment.” In *Mechanisms of Activity and Unrest at Large Calderas*, ed. Claudia Troise, Giuseppe De Natale, and Christopher R. J. Kilburn, 141–57. London: Geological Society.
- Bishop, Libby, and Arja Kuula-Luomi. 2017. “Revisiting Qualitative Data Reuse: A Decade On.” *SAGE Open* 7 (1).
- Choi, Charles Q. 2017. “Famous 600-Year-Old Nova Pinpointed in Modern Day.” <https://www.space.com/37985-600-year-old-nova-pinpointed-modern-day.html>.
- Chuine, Isabelle, Pascal Yiou, Nicolas Viovy, Bernard Seguin, Valérie Daux, and Emmanuel Le Roy Ladurie. 2004. “Historical Phenology: Grape Ripening as a Past Climate Indicator.” *Nature* 432:289–90.
- Cook, Alan. 1998. *Edmond Halley: Charting the Heavens and the Seas*. Oxford: Clarendon.
- Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. “Science Friction: Data, Metadata, and Collaboration.” *Social Studies of Science* 41:667–90.
- Forero-Mendoza, Sabine. 2002. *Le temps des ruines: Le goût des ruines et les formes de la conscience historique à la Renaissance*. Seyssel: Champ Vallon.
- Griffin, R. Elizabeth. 2006. “Detection and Measurement of Total Ozone from Stellar Spectra: Paper 2. Historic Data from 1935–1942.” *Atmospheric Chemistry and Physics* 6:2231–40.
- . 2017. “Rescue Old Data Before It’s Too Late.” *Nature* 545:267.
- Griffin, R. Elizabeth, and the CODATA Task Group Data at Risk. 2015. “When Are Old Data New Data?” *GeoResJournal* 6:92–97.
- Grindlay, Jonathan E., Sumin Tang, Robert Simcoe, Silas Laycock, Edward Los, D. J. Mink, Alison Doane, and George Champine. 2009. “DASCH to Measure (and Preserve) the Harvard Plates: Opening the ~100-Year Time Domain Astronomy Window.” In *Preserving Astronomy’s Photographic Legacy: Current State and the Future of North American Astronomical Plates*, ed. Wayne Osborn and Lee Robbins, 101–10. San Francisco: Astronomical Society of the Pacific.
- Guidoboni, Emanuela. 2010. “History and Volcanology: Dialogue Overdue? The Case of a False Eruption in Medieval Italy.” *Eos* 91:231–32.

- Guidoboni, Emanuela, and Cecilia Ciuccarelli. 2011. "The *Campi Flegrei* Caldera: Historical Revision and New Data on Seismic Crises, Bradysseisms, the *Monte Nuovo* Eruption and Ensuing Earthquakes (Twelfth Century–1582 AD)." *Bulletin of Volcanology* 73:655–77.
- Hammersley, Martyn. 1997. "Qualitative Data Archiving: Some Reflections on Its Prospects and Problems." *Sociology* 31:131–42.
- . 2010. "Can We Re-use Qualitative Data via Secondary Analysis?" *Sociological Research Online* 15 (1): 5.
- Hanson, Norwood R. 1958. *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press.
- Hayakawa, Hisashi, Harufumi Tamazawa, Yurina Uchiyama, Yusuke Ebihara, Hiroko Miyahara, Shunsuke Kosaka, Kiyomi Iwashashi, and Hiroaki Isobe. 2017. "Historical Auroras in the 990s: Evidence for Great Magnetic Storms." *Solar Physics* 292:12.
- Heaton, Janet. 2004. *Reworking Qualitative Data*. London: Sage.
- Hedenstierna-Jonson, Charlotte, Anna Kjellström, Torun Zachrisson, Maja Krzewińska, Veronica Sobrado, Neil Price, Torsten Günther, Mattias Jakobsson, Anders Götherström, and Jan Storå. 2017. "A Female Viking Warrior Confirmed by Genomics." *American Journal of Physical Anthropology* 164:853–60.
- Hsia, Florence. 2017. "Astronomy after the Deluge." In *Science in the Archives: Pasts, Presents, Futures*, ed. Lorraine Daston, 17–52. Chicago: University of Chicago Press.
- Jesch, Judith. 2017. "Let's Debate Female Viking Warriors Yet Again." *Norse and Viking Ramblings*, September 9. <http://norseandviking.blogspot.nl/2017/09/lets-debate-female-viking-warriors-yet.html>.
- Jones, Phil. 2008. "Historical Climatology—a State of the Art Review." *Weather* 63:181–86.
- Kronk, Gary W. 1999. *Cometography: A Catalog of Comets*. Vol. 1, *Ancient–1799*. Cambridge: Cambridge University Press.
- Law, John. 2004. *After Method: Mess in Social Science Research*. London: Routledge.
- Leonelli, Sabina. 2018. "The Time of Data: Timescales of Data Use in the Life Sciences." *Philosophy of Science*, in this issue.
- Le Roy Ladurie, Emmanuel. 1959. "Histoire et climat." *Annales: Economies, Sociétés, Civilisations* 14:3–34.
- Mauthner, Natasha S., Odette Parry, and Kathryn Backett-Milburn. 1998. "The Data Are Out There, or Are They? Implications for Archiving and Revisiting Qualitative Data." *Sociology* 32:733–45.
- Moore, Niamh. 2007. "(Re)using Qualitative Data?" *Sociological Research Online* 12 (3).
- Pyle, David M. 2017. *Volcanoes: Encounters through the Ages*. Oxford: Bodleian Library.
- Rabinow, Paul, and William M. Sullivan, eds. 1987. *Interpretive Social Science: A Second Look*. Berkeley: University of California Press.
- Robbirt, Karen M., Anthony J. Davy, Michael J. Hutchings, and David L. Roberts. 2011. "Validation of Biological Collections as a Source of Phenological Data for Use in Climate Change Studies: A Case Study with the Orchid *Ophrys sphegodes*." *Journal of Ecology* 99:235–41.
- Rudwick, Martin J. S. 2008. *Worlds before Adam: The Reconstruction of Geohistory in the Age of Reform*. Chicago: University of Chicago Press.
- Schechner, Sara J., and David Sliski. 2016. "The Scientific and Historical Value of Annotations on Astronomical Photographic Plates." *Journal for the History of Astronomy* 47:3–29.
- Shara, Michael M., et al. 2017. "Proper-Motion Age Dating of the Progeny of Nova Scorpii AD 1437." *Nature* 548:558–60.
- Sharma, Sapna, John J. Magnuson, Ryan D. Batt, Luke A. Winslow, Johanna Korhonen, and Yasuyuki Aono. 2016. "Direct Observations of Ice Seasonality Reveal Changes in Climate over the Past 320–570 Years." *Scientific Reports* 6:25061.
- Slavnić, Zoran. 2013. "Towards Qualitative Data Preservation and Re-use—Policy Trends and Academic Controversies in UK and Sweden." *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 14 (2): 10.
- Stanley, Matthew. 2013. "Where Is That Moon, Anyway? The Problem of Interpreting Historical Solar Eclipse Observations." In *"Raw Data" Is an Oxymoron*, ed. Lisa Gitelman, 77–88. Cambridge, MA: MIT Press.
- Stephenson, F. Richard, and David A. Green. 2002. *Historical Supernovae and Their Remnants*. Oxford: Clarendon.

- . 2005. "A Reappraisal of Some Proposed Historical Supernovae." *Journal for the History of Astronomy* 36:217–29.
- Tarrant, Anna. 2017. "Getting Out of the Swamp? Methodological Reflections on Using Qualitative Secondary Analysis to Develop Research Design." *International Journal of Social Research Methodology* 20:599–611.
- Thompson, Paul. 2000. "Re-using Qualitative Research Data: A Personal Account." *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 1 (3): 27.
- Vanderbilt, Kristin, and David Blankman. 2017. "Reliable Metadata and the Creation of Trustworthy, Reproducible, and Re-usable Data Sets." In *Stepping in the Same River Twice: Replication in Biological Research*, ed. Ayelet Shavit and Aaron M. Ellison, 179–94. New Haven, CT: Yale University Press.
- Wilkinson, David T., and P. J. E. Peebles. 2000. "Discovery of the Cosmic Microwave Background." *Physica Scripta* T85:136–41.
- Willis, Charles G., et al. 2017. "Old Plants, New Tricks: Phenological Research Using Herbarium Specimens." *Trends in Ecology and Evolution* 32:531–46.
- Wylie, Alison. 2017. "How Archaeological Evidence Bites Back: Strategies for Putting Old Data to Work in New Ways." *Science, Technology, and Human Values* 42:203–25.