

## CHAPTER 5\*

# Overlooked and Overrated Data Sharing

## Why Some Scientists Are Confused and/or Dismissive

*Heidi J. Imker*

Data curation, particularly within academic libraries, has gained appreciable momentum by developing an energetic community dedicated to providing widespread access to well-curated data. In one vision of the future, the data required to validate or extend a research study is readily available, and the publication of data itself will bear an equal importance to that of the article publication. The data curation community is eager to help catalyze that transformation through services and advocacy. Yet in practice, it's not uncommon to encounter scientists who question the cost-benefit ratio of the time and effort involved with curation, publication, and preservation of research data. How can something that seems so self-evident to the data curation community be so challenging to implement in the wild?

One possible reason is that libraries and the data curation community gravitate towards the progressive ideals of open science;<sup>1</sup> however, by its very nature, progressive is not representative. Data curators are well acquainted with the shortcomings of current data sharing practices, such as the over-use of PDFs for data publication, which restricts reuse by encapsulating otherwise useful data in this traditional publication format. However, such practices have been in place

---

\* This work is licensed under a Creative Commons Attribution 4.0 License, CC BY (<https://creativecommons.org/licenses/by/4.0/>).

for decades, and frustration with those practices is not uniform; there is rarely one voice that emerges from a given community of practice, let alone unification across all research communities.<sup>2</sup> The aim of this chapter is to take a fresh look at current practices and the nuances that surround data sharing in order to hone our messages and services as data curators with a range of perspectives in mind.

This chapter will first contextualize data sharing in the United States by looking at cultural expectations and norms within science communities. We'll then examine how scientists have historically shared research data, particularly long before modern public access requirements, since this is a useful way to frame current practice. Overlooking presently active, albeit seemingly imperfect, forms of data sharing, while ignoring researchers' own experiences and perspectives, can lead to confused or dismissive reactions to data sharing mandates and outreach. Understanding this challenge is key for those in the data curation community who are attempting to garner researcher buy-in for resources and services in support data sharing activities. In particular, some forms of sharing are successful and worthy of reexamination in light of their prevalence and adoption, even if they involve methods that do not meet data curation community approval. Finally, several large-scale data sharing efforts have been unsuccessful, and examination of the circumstances that led to their sun-setting is informative as well. The data curation community is understandably receptive to the issues that drive increased data sharing, namely transparency, reuse, and reproducibility, but we must also acknowledge the limitations of data sharing for the healthy and sustainable development of the data curation field.

## Data Sharing in Context

As funding shrinks and expectations expand, it is not surprising that researchers consistently list time, cost, and appropriateness (such as sensitivity, confidentiality, or IP protection) as barriers to data sharing.<sup>3</sup> In 2005, the administrative burdens required to execute federally funded research became so overwhelming and problematic that the topic escalated to large-scale review by the Federal Demonstration Partnership.<sup>4</sup> Despite some efforts to reform and streamline reporting activities over the following decade, only 57.7 percent of faculty's available research time was actually spent on active research.<sup>5</sup> The rest of the time was spent on administrative tasks for research, largely preparing new proposals and reporting on awarded grants. While data-sharing efforts could be considered as part of active research, it cannot be ignored that the time available for *all* aspects of active research is limited.

The need for extramural funding in the sciences feeds directly into the time shortage mentioned above. The percentage of US grants submitted that are successfully awarded, known as "funding success," has steadily decreased in recent

years, from roughly 1 in 3 being awarded in 2001 to roughly 1 in 5 being awarded in 2013.<sup>6</sup> Reduction in funding success can be attributed to many causes, but both increased demand (i.e., more grant proposals submitted) and less federal funding when adjusted for inflation are prominent reasons.<sup>7</sup> Loss of grant funding in the sciences, especially over extended periods of time, results in the inability to fund material purchase, equipment allocation, and graduate student, postdoc, or staff salaries. This dramatically slows project progression, including publication activity, and reduces subsequent competitiveness on future applications. For example, when an investor submits an application to renew an NIH grant, the review panel “will consider the progress made in the last funding period,” and the criteria include demonstration of “an ongoing record of accomplishments that have advanced their field(s).”<sup>8</sup> When Tenopir’s 2015 follow-up survey on data-sharing practices and perceptions included “I need to publish first” as a potential barrier, it was rendered the new top concern through affirmation by 43.5 percent of respondents.<sup>9</sup>

Grants are also a source of support for institutions through recovery of operating costs (e.g., administrative support, operation and maintenance of physical space, etc.). Recovery occurs by application of an “indirect cost rate” to funds awarded, and the rate is derived through a negotiation between the grantee institution and funding organization. Rates may vary dramatically, but for illustrative purposes we can use an average rate of 58.2 percent based on 49 institutional rates recently compiled.<sup>10</sup> In the most straightforward scenario, if an investigator is awarded \$100,000 in direct costs for a project, an additional \$58,200 is provided to the institution for indirect costs, resulting in a total award of \$158,200 from the funding agency. Thus fewer grants mean less funding not just for the investigator, but also for the institution.

With productivity hampered and financial pressure at the institution, loss of funding for a faculty member may come with marginalization within the scientific community and within the institution. Marginalization at the institution may result in reduction in lab space, increased teaching or administrative load, or lack of input into decisions. Tenured faculty are by no means immune to marginalization, but a lack of funding for untenured faculty places them at a distinct disadvantage. As a result, in the sciences pretenure faculty are urged to focus on securing external grants as a requirement for promotion.<sup>11</sup> While cultural changes for openness and sharing may be occurring, the reality for today—and most likely for several years to come—is that the average academic scientist will focus his or her finite time on what ensures continued funding and job security. And as data curators we must think strategically to work within this reality.

Therefore, as we consider data curation work, it’s important to keep in mind that a single definition of what constitutes data sharing cannot be extrapolated across all domains, since scientific disciplines themselves have the latitude to define what data means within each of their disciplines.<sup>12</sup> In fact, even within

domains, data sharing takes on an a myriad of forms; for example, the U.S. Geological Survey Manual states “USGS scientific data may be released or disseminated in a variety of ways, for example in datasets and databases, software, and other information products including USGS series publications (SM 1100.3), outside publications (SM 1100.4), and USGS Web pages.”<sup>13</sup> This sort of cultural relativism may be a frustration within the data curation community since it could possibly enable data withholding. However, disciplines are grappling with the current ambiguity of “data” itself,<sup>14</sup> let alone “data sharing.”<sup>15</sup> This isn’t entirely surprising. During examination of a similar semantic data topic, Renear, Sacchi, and Wickett stated that while a precise definition of *dataset* is desirable to the data curation community, informational definitions are generally functional and specific to a given discipline.<sup>16</sup> Efforts to define data sharing on behalf of a community are likely to be dismissed, and by talking at cross-purposes, data curators may lose the opportunity to nurture the evolution of those definitions within scientific communities.

While the data curation community often focuses on scientists not sharing research data, evidence that scientists do share data is prolific. Many reports, including surveys, case studies, and even data-withholding studies, indicate successful data sharing does exist. For example, surveys of researcher data-sharing practices consistently report that researchers do share their data. In 2011 Tenopir and colleagues found that only 9.6 percent of respondents somewhat or strongly disagreed with the statement “I share my data with others,” whereas the vast majority of respondents, 74.9 percent, strongly or somewhat agreed with the statement; the majority believed that they were sharing data at least to some extent.<sup>17</sup> Moreover, Tenopir and colleagues found that this sentiment increased in the 2015 follow-up study.<sup>18</sup> The 2014 Wiley study on data sharing found that 36 percent to 66 percent of researchers across five major disciplines self-reported sharing their data.<sup>19</sup> Within this study, the highest reported reason for hesitancy was intellectual property or confidentiality issues, both of which are well-acknowledged exceptions, even within the OSTP memo itself.<sup>20</sup> These concerns may account for the discipline reporting the lowest sharing: social scientists; however, openly shared data for the Wiley survey is ironically not yet available. A few empirical studies of data withholding have shown less data sharing in practice than the self-reporting survey results, albeit several of these studies have also been in disciplines that involve human subject research and therefore are more likely to be subject to ethical concerns.<sup>21</sup> Regardless of sensitivities, the results did not conclude that zero data sharing occurred. Examination of articles postpublication for evidence of shared data also revealed that sharing routinely occurs in practice and is not just an unsought ideal.<sup>22</sup>

Although not the focus of this chapter, it’s important to note the seemingly conflicting messages being directed at researchers regarding sensitive data. In particular, the rigorous procedures required for protecting human subjects car-

ry serious ramifications if breached, and researchers are constantly reminded of their obligations.<sup>23</sup> Furthermore, when the White House announced policies for government-generated Open Data, it also warned of individual identification through the “mosaic effect,” which occurs when nonidentifying data is combined with other available data to enable identification.<sup>24</sup> While a supposed fear of inappropriate disclosure could be used as a crutch to avoid data sharing, in this complex environment one person’s data withholding may be another person’s genuine concern about data breach or lack of adequate informed consent. Social and behavioral sciences have developed methodologies, protocols, and systems to allow appropriate dissemination of some restricted-use data, and repositories such as ICPSR offer excellent resources and guidance.<sup>25</sup> Thoughtful implementation of appropriate procedures and practices must be crafted at the point of project conceptualization such that the results are ultimately useful to the research community but also safe and ethical for participants. Through proactive engagement with researchers, data curators can be the gateway to such information before a study even begins and therefore increase the likelihood that study design will enable future data sharing.

Given this environment, how have scientists traditionally shared data? The next sections of this chapter will explore several overlooked ways in which researchers may already be sharing their data.

## *Overlooked Data Sharing: Article Publication*

Scientists frequently think of article publication as a form of data sharing, and it is critical to acknowledge not only that this concept exists, but also that it has been recapitulated throughout their communities, including funding agencies. As of October 2015, NIH’s Data Sharing workbook still says “Some studies, such as small laboratory-based projects, make raw data available in publications.”<sup>26</sup> Likewise, example data management guidance available from NSF and USGS websites reference data sharing via publication.<sup>27</sup> In an analysis of 1,260 Data Management Plans (DMPs) submitted for NSF applications at the University of Illinois, Mischo, Schlembach, and O’Donnell found “publication” listed as a data sharing mechanism 44 percent of the time.<sup>28</sup> So herein lies an important cultural disconnect in data sharing: as data curators, we are overlooking what many in scientific communities believe is an acceptable form of data sharing because it doesn’t fit into *our* definition of data sharing. It cannot be overemphasized that what may be substandard for the data curation community does not trump what is standard for a community of practice; *cultural norms are a critical driver for practice.*<sup>29</sup>

As an analogy, let's consider an example where someone uploads a presentation to a web service, where the slide deck is saved as a PDF without comments, animation, or the ability to manipulate. While it's obvious the slides could be shared in a manner more amenable to reuse by providing the original presentation format, could one say that the person who posted slides via PDF did not share because the format precludes ready reuse? Is sharing in this context really a true-or-false question? It might be worth reinforcing that accuracy is fundamental to science, and therefore the question of data sharing itself is confusing when presented through application of Boolean logic, with binary true/false variables. Fuzzy logic, with many-valued variables, is more appropriate. For that reason, our messages to scientists must emphasize *how* data is shared as opposed to the singular act of data sharing itself. Amending and clarifying our language by using phrases such as "reuse-ready sharing," "fit-for-purpose sharing," or "source file sharing" is one step in that direction.

Consider a recent study from Ron Vale that examined the amount of data shared through publication by comparing figures in publications in the journals *Nature*, *Cell*, and the *Journal of Biological Chemistry* for years 1984 and 2014.<sup>30</sup> Figures are a critical component of academic work and can present data (including raw, aggregate, and representative) through tables, graphs, images, schematics, and more. Through scoring of figures and panels, Vale concluded that publications include 2 to 4 times more data in 2014 than they did in 1984. The increase in data-per-publication ties into time-to-publication, which has slowed according to Vale's analysis. He attributed both trends largely to the need to publish comprehensive studies that provide an exhaustive and, especially in the eyes of the reviewer, hopefully unequivocal argument that the findings are valid. This sentiment has been echoed elsewhere during interviews with scientists.<sup>31</sup> Interestingly, Vale expressed frustration at the amount of data acquisition that is required for such "mature" studies. He noted that while some reviewer suggestions improve the work, "many suggested experiments [that] are unnecessary, and sometimes the requested work is so extensive that it constitutes a separate study onto itself."<sup>32</sup>

Vale's article preprint posted to bioRxiv.org resonated well within the scientific community by garnering thousands of views and hundreds of social media hits, and it was later published with peer review.<sup>33</sup> Vale's ultimate argument was for faster publication, particularly through publication of smaller studies and use of preprint servers. These solutions are consistent with the open science values of the data curation community. Pragmatically, more publication of "partial" studies would also likely yield smaller, more readily curated data sets; quicker time to data sharing could likely curb some information entropy. Nonetheless, the potential synergy could be wasted unless there is an effort to understand that the resistance to greater data sharing may have a deeper-seated resistance that is rooted in the broader data-related demands placed on researchers during other

parts of the research process. Our message has to be laser-focused on the value of data set availability and curation, and not focused simply on “data sharing” since so many within the scientific communities view article publication as data sharing already. Without addressing this critical nuance we may become just another voice seeming to arbitrarily demand that more time and effort be spent on data.

## *Overlooked Data Sharing: Supplemental Material*

Similarly, a form of data sharing often overlooked in the data curation community is supplemental material provided along with a published journal article, also known as supplemental data, auxiliary information, supporting information, or supplementary content.<sup>34</sup> Supplemental material is generally supplied to a publication in free form as an extra file or files that help support the main article. A prototypical example is a PDF that may include additional text, methods, analyses, figures, tables, and/or data, but other supplement examples may include file formats incompatible with article format or layout, such as video or code.<sup>35</sup> *PLOS* and *Science*, as just two examples, allow a myriad of file formats as supplemental files.

There are several reasons for providing supplemental material, such as allowing a reader to focus on the most salient points in the main body of the article or allowing the reader to access material that logistically cannot be placed within the main body due to size or format. Authors may submit supplements as a way to demonstrate that their work is thorough and well-executed, or they may submit under the belief that extra material may help “immunize” them from reviewer concerns.<sup>36</sup> Material that may have belonged in the main body is sometimes otherwise relegated to supplemental files due to journal space considerations or to minimize author page costs.<sup>37</sup>

Supplemental material is often tied to the advent of the electronic journal, but scientists have been providing more detail for primary articles via supplements for decades (for early examples in print see Myers and Abeles in 1990 and Sapp, Lord, and Hammarlund in 1975).<sup>38</sup> However, rapid adoption of electronic supplemental materials began in the late 1990s.<sup>39</sup> Over the course of a decade, Beauchamp of *The Journal of Clinical Investigation* reported that the percent of articles containing a supplement jumped from just 3 percent in 2001 to 95 percent in 2011.<sup>40</sup> Similar results have been reported for *The Journal of Experimental Medicine* and *The Journal of Neuroscience*.<sup>41</sup> Kenyon and Sprague’s thorough analysis of sixty journals broadly covering the environmental sciences similarly found that supplemental file adoption picked up quickly, albeit not entirely uniformly, between 2000 and 2011.<sup>42</sup>

The rapid adoption of supplemental materials suggests a successful data sharing mechanism; however, the practice has not been without debate. Libraries are concerned with the apparent “Pandora’s box of management issues” including a lack of document structure, metadata, persistence, and discoverability.<sup>43</sup> Journal editors have also weighed in with their concerns about quality, overhead, and the relevance of supplemental materials.<sup>44</sup> At least one journal has banned supplemental materials altogether over these concerns,<sup>45</sup> and others have implemented policies that limit supplemental materials to only that which is “essential.” On the other hand, many journals encourage supplements and also recommend a variety of file formats beyond just PDFs (e.g., see table 2 in Kenyon and Sprague).<sup>46</sup> In light of these inconsistencies and concerns, NISO and NFAIS established a formal working group in 2010 to develop recommended practices.<sup>47</sup> This group uncovered a messy landscape both in opinion and practice. Not only is the content highly variable, the handling of supplemental material by journals is idiosyncratic as well. For example, sometimes supplements are peer-reviewed, sometimes not; sometimes supplements and articles are formally linked, sometimes not. Culturally, Swartzman found two distinct camps: those who encouraged as much additional detail as deemed necessary, and those who felt supplemental materials were being used as a “data dump.”<sup>48</sup> Although one might be inclined to dismiss the concerns of journal editors as business-motivated rather than scientific-value-motivated, this is not the only arena where “overflow” concerns have emerged. During interviews with biomedical researchers, Siebert, Machesky, and Insall found that interviewees expressed many overflow-related concerns, including the proliferation of new journals, the explosion of publications, and even an excess of scientists themselves. This cumulated in an overarching concern that “rapid proliferation of scientific outputs was inconsistent with the capacity of the world of science to verify the quality of outputs.”<sup>49</sup>

Regardless of the greater scientific community’s ability to process the deluge of information, it’s clear that many scientists *are* willing to share additional information via supplements, and at least some portion of the scientific community appreciate the added content. Although supplemental materials may contain more than data, data curators’ skills squarely align with addressing the flaws of supplemental materials: unstructured information, lack of metadata, uncertain access persistence, and limited discoverability. Indeed, “Most frequently, supplemental materials suffer from a lack of descriptive metadata.”<sup>50</sup> As data curators we can view supplemental materials as a positive model and can pitch curation services as being able to alleviate several of the drawbacks that vex research communities. For communities that have embraced supplemental materials, one model may be able to encourage researchers to think of deposit into data repositories as “Upgraded Supplemental Materials,” where upgrade may mean something along a continuum of minimal metadata at one end to detailed curation at the other, depending on the scope of services available. Here we can emphasize consistent

metadata, persistent identifiers, stability, availability, and file format flexibility as directly addressing the nearly universally acknowledged limitations of supplemental files. While not perfect, even the most minimal, unmediated deposit is a huge step towards progress when compared to the current haphazard landscape of supplemental materials as described here.

## *Overrated Data Sharing: Unsustained Community Resources*

In juxtaposition to the unstructured nature of supplemental materials or the limitations of published articles, an untold number of highly structured and sophisticated data resources have also been developed. When the topic of domain repositories is broached, successful well-known examples such as Inter-university Consortium for Political and Social Research (ICPSR), GenBank, or the Sloan Digital Sky Survey quickly spring to mind; however, as of October 2015 the Registry of Research Data Repositories, re3data.org, contained 1,363 reviewed repositories with representation across both the humanities and sciences.<sup>51</sup> The number of resources represented in re3data.org is steadily growing, and it's understood that the registry is not yet comprehensive. For example, since 1993 *Nucleic Acids Research* has published an annual "Database Issue" and maintained an online Molecular Biology Database Collection that currently references 1,549 databases dedicated solely to bioinformatics and molecular biology.<sup>52</sup> Thus it's difficult to estimate how many data resources are currently available, but clearly data resources are of keen interest to many research communities.

Sustaining resources, however, is a much different animal. Established repositories are often asked to absorb endangered data, as recently occurred when the Cultural Policy and the Arts National Data Archive (CPANDA) began migration of data to the ICPSR and the National Archive of Data on Arts and Culture (NADAC) after conclusion of funding.<sup>53</sup> However, a lack of committed funding is a major concern for even the most successful and well-used domain repositories.<sup>54</sup> Recalcitrant funding agencies are extremely hesitant to commit to funding anything in perpetuity, citing their missions to spur innovation and the need to be responsive to new scientific directions. To be fair, the agencies are in a difficult position. As the number of new resources increases over time, the amount of funding required to sustain those resources would likewise accumulate. Without triage or alternative support mechanisms, undoubtedly funders fear that sustaining infrastructure will disproportionately result in reduced funding for new research.

This has created a habitual scenario where resources are left in limbo to scramble for support. In some cases, resources have been "sunsetting" due to lack

of community use or buy-in. Interestingly, in their exploration of data sharing behavior in the social sciences, Kim and Adler found that just because a data repository exists does not mean a community finds value in it.<sup>55</sup> One high profile example in the biological sciences is the Knowledgebase for the Protein Structure Initiative (PSI), a fifteen-year program funded through NIH's National Institute of General Medical Sciences, which aimed to advance technologies for the determination of three-dimensional protein structures. On conclusion of the PSI project, three review committees jointly concluded that the resource had yet to demonstrate broad use across the user communities and fate of the PSI Knowledgebase lies unknown.<sup>56</sup> Similar concerns were expressed for the recently sunsetted Virtual Astronomy Observatory.<sup>57</sup> When promises are made that such resources will empower scientific communities by providing access to data and yet the resources fail to live up to that promise, it's disillusioning to scientists who are already frustrated by the hypercompetitive funding climate. The arguments against these resources were that the money could be better spent elsewhere. Institutions with funds devoted to data curation and repositories meant to support data sharing are no less susceptible to such budgetary criticism at the local level; thus buy-in from local scientific communities is essential.

However, it's not only a lack of community buy-in that has doomed some resources. For example, in 2007 the National Library of Medicine announced plans to cut funding to five community resources and redirect funds towards "research and training." The resources had several thousand users, and communities attempted to rally in order to save them.<sup>58</sup> Likewise, the extremely popular Kyoto Encyclopedia of Genes and Genomes (KEGG) issued pleas in 2011 reminiscent of a National Public Radio pledge drive after restructuring of the primary Japanese funder.<sup>59</sup> Users who benefited from KEGG were urged "to write, email, tweet, and blog about your support for KEGG. I hope, in the long run, your voices will increase our chances of getting more stable funding."<sup>60</sup> KEGG has turned to a partly commercial model, but is still not fully sustainable. Time and again, resources have been put in peril despite demonstrated value to communities.

Notwithstanding the clear inability to sustain each new resource developed, researchers have had a penchant for developing such resources, frequently as a by-product of a larger research project (such as the PSI Knowledgebase as part of the larger PSI program described above). Likewise, funding agencies have a penchant for enabling such efforts, if not outright encouraging or requiring them. On one hand, these resources stand as further testaments to active data sharing. On the other hand, post-grant support planning has not been emphasized until recently, as evidenced by adoption of data management plans by federal grant agencies, and even today there has been no dissuasion of standing up isolated resources that will ultimately need migration, rescue, or sunseting. This has created a culture of at-risk data with no end in sight. These high-profile failures—

whether they represent lack of community use or lack of sustained funding—are another reason why scientists may be doubtful since they cast a shadow of hopelessness on data sharing. Indeed, such efforts begin to look overrated. One critical thing we can do as data curators is attempt to circumvent diversion of funds into one-off resources and instead emphasize the importance of centralized, community-based solutions whether they include our local institutional repositories or domain data repositories housed at other institutions. A major hurdle for us to achieve this will be aligning idiosyncratic needs of unique projects with the broad service models of community resources. Here, we can remind researchers that giving up the customization and control of a uniquely developed resource allows for more project funds and energy to go to the research at hand.

## *Overrated Data Sharing: Hyperbolic Arguments*

It is not a foregone conclusion that all data, even that without restrictions, should be shared. All data is not equally valuable, and several public access implementation plans have made it clear that they do not expect all data to be available. For example, the NIH states, “It is important to note that not all digital scientific data need to be shared and preserved.”<sup>61</sup> Likewise, NSF plan stated, “rarely does NSF expect that retention of all data that are streamed from an instrument or created in the course of an experiment or survey will be required.”<sup>62</sup> In fact, the OSTP memo itself expects that agency plans will take into account “preserving the balance between the relative value of long-term preservation and access and the associated cost and administrative burden.”<sup>63</sup>

Not only is the data not always required to validate or reproduce research results, but the reuse utility varies dramatically between discipline, purpose of original study, and data types (e.g., see Borgman’s 2012 discussion of data types categorized as observational, computational, experimental, and records).<sup>64</sup> There is no universal approach, and broad data availability is not yet mature enough for ready identification of data that has enduring value. Furthermore, as Borgman noted, “Perhaps the most significant challenge to data sharing is the lack of demonstrated demand for research data outside of genomics, climate science, astronomy, social science surveys, and a few other areas.”<sup>65</sup> This is a reality that dramatically complicates the data-sharing landscape. Efforts such as the Stewardship Gap Project aim to clarify this reality by developing evaluation frameworks and recommendations to identify data of particularly high value along with the support required to ensure long-term access.<sup>66</sup> Because of the current ambiguity, however, overemphasis on the impact of data specifically may also confuse or even aggravate some researchers.

Rather than reuse the data, some data will simply be replaced because the original data has only transitory use for a specific experiment. Take an example of observing the growth of a bacteria population. One way to measure growth is to inoculate a liquid culture with a very small amount of “starter” bacteria from a pure stock of bacteria. The liquid starts out clear, and the researcher essentially measures the increase in “cloudiness” of the liquid as the bacteria grow over time. The raw data is a series of time points and the density (“cloudiness”) measurement at each of those times, which is then represented as a graph of density ( $y$ -axis) versus time ( $x$ -axis). If the wrong bacterial stock was mistakenly used to conduct the growth experiment, neither the raw data nor the graphical representation necessarily divulge that error since it’s just a measure of bacterial density and not of bacterial type. While for some types of research, access to raw data in its original format may be helpful or even imperative, this is an example where the underlying data likely holds no more utility than representations of the data. Accounting for error and fluctuation is why independent replication *within a given study* is critical and considered a cornerstone in experimental sciences.<sup>67</sup> Should other researchers want to replicate the initial findings, they would never reuse the raw data by replotting the graph of growth. They would redo the entire experiment and acquire their own growth measurements independently to account for potential flaws or idiosyncrasies in the researcher’s execution, protocol, materials, or environment. It’s not a matter of trust in the data; it’s a matter of external verification of the experiment as a whole. In fact, Crotty and commenters argue that clear and accurate methodology is more important than data access.<sup>68</sup> On the other hand, the very same project may include a genomic analysis of the bacterial culture, and the ensuing genomic sequences may be of reuse utility. Unfortunately, because no absolutes apply, we simply cannot state that data sharing practices *are* appropriate for one data type or *are not* appropriate for another data type, even within a given discipline. It is maddeningly messy.

While scientific communities, agencies, and publishers struggle to establish which data to share or not share, scientists may feel obligated to share everything, regardless of value, which evokes the “data dump” concern already associated with supplemental materials. While perhaps overcompensation is an enviable problem, the issue of long-term value will be further exacerbated by the continued lack of definitions, standards, and best practices, which are all equally important but even more difficult to address. If some scientists share not because they—or anyone else—*truly value the data* but simply because they view data sharing as insulation against criticism or as a requirement for compliance, we have to prepare ourselves in the data curation community to ask: does this data also warrant the substantial effort of curation and preservation? We must view scientists, both as consumers and producers of data, as our best partners in determining which data should benefit from our resources and for how long.

Because the data itself is just one component of research, a single-minded focus on data can ultimately detract from increased transparency and reproducibility. Without robust experimental design, such as use of proper controls and sampling procedures, raw data may be just as erroneous as a representative figure. Likewise, simulation data is critically dependent on the software versions used, the initial parameters used in a simulation run, and the general operating variables. If data sharing alone were to become a sort of rubber stamp for better research, large swaths of science will fail in this assessment. For these reasons, all disciplines have not necessarily taken the same path towards data sharing. In 2015, NIH issued plans to enhance rigor and transparency through four major areas: (1) the scientific premise of the proposed research, (2) rigorous experimental design for robust and unbiased results, (3) consideration of relevant biological variables, and (4) authentication of key biological and/or chemical resources.<sup>69</sup> Although NIH acknowledges that data is important, clearly it is not an all-encompassing solution. In this regard when the ultimate goal is to enable better science, then the best scenario is to enable inclusion of whatever has been missing, whether that be data, code, methodology, materials, or any other information. While in some cases the term *data* has become a bucket for anything research-related that's not a journal article, acknowledging semantic differences is important for the sake of productive communication and grittier issues like the application of intellectual property law. As mentioned above in the supplemental materials section, data in the "factual material" sense is not the only thing that could benefit from best practices, standardization, and curation. While this could be a potential complication to data curation services, data curators do not necessarily have to play an active role in hands-on curation of all things research-related, especially in the short term. Simply being knowledgeable of current and emerging trends, such as new policies and new sharing platforms, is of value. Indeed, such a role aligns with the reference services that stand as a fundamental mission of libraries. The benefit of thinking more broadly will be useful in the long term to the data curation profession, however, because accumulated knowledge through such conversations will enable user-informed evolution of data curation service models.

## Conclusions

While the data curation community has been justifiably buoyed by the impact of data sharing success stories, the points presented are intended to serve as examples of the nuances that surround data sharing. As data curators, we do ourselves a disservice if we look at data sharing only from the perspective of progressive or idealist attitudes. Without attempting to understand and accommodate the nuances of data sharing, then the lack of rapid, dedicated, and widespread adoption

of new practices will lead to frustration in the data curation community. Indeed, some antagonistic views, such as accusing scientists of misconduct, laziness, or lack of creativity if they fail to see a need for data sharing, have already surfaced in the back channels of the data curation community (e.g., social media, Listservs, and conferences),\* which may be a manifestation of frustration. Instead of setting ourselves up for disappointment, a more nimble approach is to acknowledge a broader perspective that stems from the variability of definitions, communities, practices, and science itself. For those who interface directly with scientists, ultimately our greatest effectiveness will come by virtue of working within the realities that scientists experience.

For example, the author received an e-mail some months ago from a faculty member who inquired if university-wide data sharing practices had been established. A publisher was requesting that individual-level data be made available, but the faculty member was reluctant to share. In the e-mail, the researcher initially cited the need to do a secondary analysis, the limitations of the data set, and the desire to share the data within the specific research community (as opposed to untargeted sharing) as reasons for not wanting to share openly. At first pass, some data sharing advocates would not find any of these reasons “valid.” A colleague and I met with the faculty member and two graduate students also on the project, and we devoted our time to simply listening and learning about their concerns. We learned that the publisher’s data sharing policies had changed mid-peer-review, and the faculty member held deep reservations about whether publishers, who may not be as attuned to data utility or as thoughtful of sharing consequences, are appropriate drivers of data sharing practices. We also learned that human subject participants had signed consents that stated data would be shared only in aggregate, which would mean time-consuming and potentially impossible re-consent of each participant prior to sharing deidentified participant-level data. Furthermore, if data was published from the study, the lack of accompanying control data would dramatically reduce utility. Perhaps most interestingly, we also learned that this research area had already established a committee to define best practices for data analysis and sharing, in which the faculty member participated, and a recommendations report was currently under community review. In truth, we found that faculty member was a supporter of data sharing, but felt strongly that sharing at all costs was senseless. Indeed, it

---

\* For example, at the 2016 International Digital Curation Conference, a keynote address described supplemental files as “malpractice” (Barend Mons, “Open Science as a Social Machine: Where (the...) Are the Data?” [keynote address, International Digital Curation Conference, Amsterdam, the Netherlands, February 22–25, 2016], <http://www.dcc.ac.uk/sites/default/files/documents/IDCC16/Keynotes/Barend%20Mons.pdf>), and “data whining” emerged on Twitter during one panel, for example “Lots of talk at this #IDCC16 Panel session on data whining (instead of data mining). All the reasons why people can’t share their data...” (from #IDCC16 hashtag archive at <http://bit.ly/1RsVJzt> via @alastairdunning).

was also our conclusion that the cost-benefit ratio of sharing in this case was unfavorable, and we recommended the faculty member request an exception from the editor, which ultimately proved successful. The data was not shared. Had we taken the view that unwavering promotion of data sharing is the only acceptable position, it's likely that we would have failed in establishing ourselves as credible resource. Instead, we gained the faculty member's confidence as balanced and knowledgeable professionals who are supportive of research as a whole. Notably, through our interactions the group has now adopted language for participant consent that will allow for more facile and permissive data sharing in the future.

While we must keep in mind that current practices are not uniformly contested, nor is data sharing a universal panacea, it is clear that sharing will become more commonplace in coming years. There is no doubt that data curation has had—and will continue to have—an important place in science. As data sharing practices evolve, data curators have the opportunity to craft our message and services in a way that both makes sense and delivers great value to the communities we aim to serve. The strategies include (1) acknowledging cultural pressures and norms, (2) providing directness and clarity in messaging to emphasize purpose, (3) seeking to augment or enhance current practices, and (4) embracing and planning for complexity. While such strategies may fall short of ideals, they place data curators in a position to enable more efficient and robust science through closer alignment with research communities.

## Acknowledgments

The author would like to thank Elise Dunham, Bill Mischo, Sarah Williams, editor Lisa Johnston, and anonymous reviewers for thoughtful and critical evaluation of this chapter.

## Notes

1. Anna K. Gold, "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries," *D-Lib Magazine* 13, no. 9/10 (2007), <http://works.bepress.com/agold01/4/>.
2. Christine L. Borgman, "The Conundrum of Sharing Research Data," *Journal of the American Society for Information Science and Technology* 63, no. 6 (June 2012): 1059–78, doi:10.1002/asi.22634.
3. Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame, "Data Sharing by Scientists: Practices and Perceptions," *PLoS ONE* 6, no. 6 (2011): e21101, doi:10.1371/journal.pone.0021101; Carol Tenopir, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett, "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide," *PLoS ONE*

- 10, no. 8 (2015): e0134826, doi:10.1371/journal.pone.0134826; Liz Ferguson, “How and Why Researchers Share Data (and Why They Don’t),” *Exchanges* (blog), November 3, 2014, <https://web.archive.org/web/20160116150325/http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/>; Sarah C. Williams, “Data Sharing Interviews with Crop Sciences Faculty: Why They Share Data and How the Library Can Help,” *Issues in Science and Technology Librarianship*, Spring 2013, doi:10.5062/F4T151M8.
4. Robert S. Decker, Leslie Wimsatt, Andrea G. Trice, and Joseph A. Konstan, *A Profile of Federal-Grant Administrative Burden among Federal Demonstration Partnership Faculty*, A Report of the Faculty Standing Committee of the Federal Demonstration Partnership (Federal Demonstration Partnership, January 2007), [http://web.archive.org/web/20160214195603/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pgasite\\_054586.pdf](http://web.archive.org/web/20160214195603/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pgasite_054586.pdf).
  5. Sandra L. Schneider, Kirsten K. Ness, Sara Rockwell, Kelly Shaver, and Randy Brutkiewicz, *2012 Faculty Workload Survey*, research report (Federal Demonstration Partnership, April 2014), [http://web.archive.org/web/20151022202705/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pgasite\\_087667.pdf](http://web.archive.org/web/20151022202705/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pgasite_087667.pdf).
  6. National Institutes of Health, “Success Rates and Funding Rates. Research Project Grants: Competing Applications, Awards, and Success Rates 1998–2014,” NIH Data Book, 2015, <http://web.archive.org/web/20151022204459/http://report.nih.gov/NIDatobook/Charts/Default.aspx?showm=Y&chartId=124&catId=13>; National Science Foundation, *Report to the National Science Board on the National Science Foundation’s Merit Review Process: Fiscal Year 2013* (Arlington, VA: National Science Foundation, May 2014), <http://web.archive.org/web/20151022211742/https://www.nsf.gov/nsb/publications/2014/nsb1432.pdf>.
  7. IPAMM Working Group, *Impact of Proposal and Award Management Mechanisms* (Arlington, VA: National Science Foundation, 2007), <http://web.archive.org/web/20151024203046/http://www.nsf.gov/pubs/2007/nsf0745/nsf0745.pdf>.
  8. National Institutes of Health, “Definitions of Criteria and Considerations for Research Project Grant (RPG/X01/R01/R03/R21/R33/R34) Critiques,” National Institutes of Health Grants and Funding, last updated March 21, 2016, [http://web.archive.org/web/20160325020441/https://grants.nih.gov/grants/peer/critiques/rpg\\_D.htm](http://web.archive.org/web/20160325020441/https://grants.nih.gov/grants/peer/critiques/rpg_D.htm).
  9. Tenopir et al., “Changes in Data Sharing.”
  10. Jeremy Berg, “Indirect Cost Rate Survey,” *Data Hound* (blog), May 10, 2014, <https://web.archive.org/web/20150924200621/http://datahound.scientopia.org/2014/05/10/indirect-cost-rate-survey/>.
  11. Burroughs Wellcome Fund, “Obtaining Tenure,” Career Tools, 2014, <http://web.archive.org/web/20151024181949/http://www.bwfund.org/career-tools/obtaining-tenure>; Chelsea Wald, “Redefining Tenure at Medical Schools,” *Science Careers*, March 6, 2009, doi:10.1126/science.caredit.a0900032.
  12. John P. Holdren, “Increasing Access to the Results of Federally Funded Scientific Research,” Memorandum for the Heads of Executive Departments and Agencies, Office of Science and Technology Policy, Executive Office of the President, February 22, 2013, [http://web.archive.org/web/20160115125401/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://web.archive.org/web/20160115125401/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf).
  13. US Geological Survey, “Scientific Data Management Foundation, US Geological Survey Instructional Memorandum No. IM OSQI 2015-01, US Geological Survey Manual,

- February 19, 2015, <http://web.archive.org/web/20151031000659/http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-01.html>.
14. C. Titus Brown, "Cultural Confusions about Data: The Intertidal Zone between Two Styles of Biology," *Living in an Ivory Basement* (blog), April 2, 2015, <http://web.archive.org/web/20151003163047/http://ivory.idyll.org/blog/2015-culturally-confused-about-data.html>.
  15. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research."
  16. Allen H. Renear, Simone Sacchi, and Karen M. Wickett, "Definitions of *Dataset* in the Scientific and Technical Literature," *Proceedings of the American Society for Information Science and Technology* 47, no. 1 (2010): 1–4, doi:10.1002/meet.14504701240.
  17. Tenopir et al., "Data Sharing by Scientists."
  18. Tenopir et al., "Changes in Data Sharing."
  19. Ferguson, "How and Why Researchers Share Data."
  20. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research."
  21. Caroline J. Savage and Andrew J. Vickers, "Empirical Study of Data Sharing by Authors Publishing in PLoS Journals," *PLoS ONE* 4, no. 9 (2009): e7078, doi:10.1371/journal.pone.0007078; Jelte M. Wicherts, Denny Borsboom, Judith Kats, and Dylan Moleenaar, "The Poor Availability of Psychological Research Data for Reanalysis," *American Psychologist* 61, no. 7 (October 2006): 726–28, doi:10.1037/0003-066X.61.7.726.
  22. Williams, "Data Sharing Interviews with Crop Sciences Faculty"; Philip Herold, "Data Sharing among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications," *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1244, doi:10.7710/2162-3309.1244.
  23. Ajai R. Singh and Shakuntala A. Singh, "Ethical Obligation towards Research Subjects," *Mens Sana Monographs* 5, no. 1 (2007): 107–12, doi:10.4103/0973-1229.32153.
  24. Sylvia M. Burwell, Steven VanRoekel, Todd Park, and Dominic J. Mancini, "Open Data Policy: Managing Information as an Asset," Memorandum for the Heads of Executive Departments and Agencies, Office of Management and Budget, May 9, 2013, <https://web.archive.org/web/20151104005245/https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.
  25. Inter-university Consortium for Political and Social Research (ICPSR), "Phase 5: Preparing Data for Sharing," in *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle*, 5th ed. (Ann Arbor, MI: ICPSR, 2012), 5.
  26. National Institutes of Health. *Data Sharing Workbook* (Bethesda, MD: National Institutes of Health, last revised February 13, 2004), [http://web.archive.org/web/20151116180118/https://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_workbook.pdf](http://web.archive.org/web/20151116180118/https://grants.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf).
  27. National Science Foundation, *Data Management for NSF EHR Directorate* (Arlington, VA: National Science Foundation Education and Human Resources Directorate, March 2011), <http://web.archive.org/web/20151030234444/http://www.nsf.gov/bfa/dias/policy/dmpdocs/ehr.pdf>; US Geological Survey, "Scientific Data Management Foundation."
  28. William Mischo, Mary Schlemback, and Megan O'Donnell, "An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals," *Journal of eScience Librarianship* 3, no. 1 (2014), doi:10.7191/jeslib.2014.1060.
  29. Youngseek Kim and Melissa Adler, "Social Scientists' Data Sharing Behaviors: Investigating the Roles of Individual Motivations, Institutional Pressures, and Data Repositories," *International Journal of Information Management* 35, no. 4 (August 2015): 408–18, doi:10.1016/j.ijinfomgt.2015.04.007.

30. Ronald D. Vale, "Accelerating Scientific Publication in Biology," preprint, *bioRxiv*, September 12, 2015, doi:10.1101/022368.
31. Sabina Siebert, Laura M. Machesky, and Robert H. Insall, "Overflow in Science and Its Implications for Trust," *eLife* 4 (2015): e10825, doi:10.7554/eLife.10825.
32. Vale, "Accelerating Scientific Publication in Biology."
33. Ronald D. Vale, "Accelerating Scientific Publication in Biology," *Proceedings of the National Academy of Sciences* 112 (2015): 13,439–46, doi:10.1073/pnas.1511912112.
34. Jeremy Kenyon and Nancy R. Sprague, "Trends in the Use of Supplementary Materials in Environmental Science Journals," *Issues in Science and Technology Librarianship*, Winter 2014, doi:10.5062/F40Z717Z.
35. Alexander Schwarzman, "Supplemental Information: Who's Doing What and Why" (PowerPoint presentation, CSE 2012 Annual Meeting, Seattle, WA, May 20, 2012), [http://web.archive.org/web/20151013195224/http://www.niso.org/apps/group\\_public/document.php?document\\_id=8591&wg\\_abbrev=supptechnical](http://web.archive.org/web/20151013195224/http://www.niso.org/apps/group_public/document.php?document_id=8591&wg_abbrev=supptechnical); Linda Beebe, "Supplemental Materials for Journal Articles: NISO/NFAIS Joint Working Group," *Information Standards Quarterly* 22, no. 3 (Summer 2010): 33–37, [http://web.archive.org/web/20151227001104/http://www.niso.org/apps/group\\_public/download.php/4885/Beebe\\_SuppMatls\\_WG\\_ISQ\\_v22no3.pdf](http://web.archive.org/web/20151227001104/http://www.niso.org/apps/group_public/download.php/4885/Beebe_SuppMatls_WG_ISQ_v22no3.pdf).
36. John Maunsell, "Announcement Regarding Supplemental Material," *Journal of Neuroscience* 30 (2010): 10,599–600.
37. American Physical Society, "Supplemental Data," *APS Online Style Manual*, 2003, <http://web.archive.org/web/20160325000016/http://www.apstylemanual.org/oldmanual/parts/supplemental.htm>.
38. Robert W. Myers and Robert H. Abeles, "Conversion of 5-S-Methyl-5-Thio-D-Ribose to Methionine in *Klebsiella Pneumoniae*. Stable Isotope Incorporation Studies of the Terminal Enzymatic Reactions in the Pathway," *Journal of Biological Chemistry* 265 (1990): 16,913–21; Curtis Sapp, Michael Lord, and E. Roy Hammarlund, "Sodium Chloride Equivalents, Cryoscopic Properties, and Hemolytic Effects of Certain Medicinals in Aqueous Solution III: Supplemental Values," *Journal of Pharmaceutical Sciences* 64, no. 11 (November 1975): 1884–86, doi:10.1002/jps.2600641132.
39. Nicholas R. Anderson, Peter Tarczy-Hornoch, and Roger E. Bumgarner, "On the Persistence of Supplementary Resources in Biomedical Publications," *BMC Bioinformatics* 7 (2006): 260, doi:10.1186/1471-2105-7-260; Maunsell, "Announcement Regarding Supplemental Material."
40. Alexander Schwarzman, "Supplemental Materials Survey," *Information Standards Quarterly* 22, no. 3 (Summer 2010): 23–26, [http://web.archive.org/web/20151013194617/http://www.niso.org/apps/group\\_public/download.php/4886/Schwarzman\\_SuppMatls-Survey\\_ISQ\\_v22no3.pdf](http://web.archive.org/web/20151013194617/http://www.niso.org/apps/group_public/download.php/4886/Schwarzman_SuppMatls-Survey_ISQ_v22no3.pdf).
41. Christine Borowski, "Enough Is Enough," *Journal of Experimental Medicine* 208, no. 7 (2011): 1337, doi:10.1084/jem.20111061; Maunsell, "Announcement Regarding Supplemental Material."
42. Kenyon and Sprague, "Trends in the Use of Supplementary Materials."
43. Todd Carpenter, "Supplementary Materials: A Pandora's Box of Issues Needing Best Practices," *Against the Grain* 21 (2009): 84–85; Thomas Schaffer and Kathy M. Jackson, "The Use of Online Supplementary Material in High-Impact Scientific Journals," *Science and Technology Libraries* 25, no. 1–2 (2004): 73–85, doi:10.1300/J122v25n01\_06.
44. Emilie Marcus, "Taming Supplemental Material," *Neuron* 64, no. 1 (October 2009):

- 3, doi:10.1016/j.neuron.2009.09.046; Borowski, “Enough Is Enough”; Maunsell, “Announcement Regarding Supplemental Material.”
45. Maunsell, “Announcement Regarding Supplemental Material.”
  46. Kenyon and Sprague, “Trends in the Use of Supplementary Materials.”
  47. Beebe, “Supplemental Materials for Journal Articles.”
  48. Schwarzman, “Supplemental Information.”
  49. Siebert, Machesky, and Insall, “Overflow in Science.”
  50. Beebe, “Supplemental Materials for Journal Articles.”
  51. Heinz Pampel, Paul Vierkant, Frank Scholze, Roland Bertelmann, Maxi Kindling, Jens Klump, Hans-Jürgen Goebelbecker, et al., “Making Research Data Repositories Visible: The re3data.org Registry,” *PLoS ONE* 8, no. 11 (2013): e78080, doi:10.1371/journal.pone.0078080.
  52. Michael Y. Galperin, Daniel J. Rigden, and Xosé M. Fernández-Suárez, “The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection,” *Nucleic Acids Research* 43, no. D1 (2015): D1–5, doi:10.1093/nar/gku1241.
  53. CPANDA, “What Is CPANDA?” 2015, <http://web.archive.org/web/20151120151725/http://www.cpanda.org/cpanda/about>.
  54. Carol Ember and Robert Hanisch, “Sustaining Domain Repositories for Digital Data: A White Paper,” 2013, [http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper\\_ICPSR\\_SDRDD\\_121113.pdf](http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf).
  55. Kim and Adler, “Social Scientists’ Data Sharing Behaviors.”
  56. Future of Structural Biology Committees, *Recommendations for Continued Investment in Structural Biology Following the Sunset of the Protein Structure Initiative* (Bethesda, MD: National Institute of General Medical Sciences, December 2014), <http://web.archive.org/web/20151013183712/http://www.nigms.nih.gov/News/reports/Documents/NIGMS-FSBC-report2014.pdf>.
  57. Stephen Kent, Chryssa Kouveliotou, David Meyer, Richard H. Miller, David Schade, James Schombert, Alexander Szalay, and Suresh Santhana Vannan, *Report of the Astrophysics Archives Program Review for the Astrophysics Division, Science Mission Directorate* (NASA, May 6–8, 2015), <http://web.archive.org/web/20151107185033/http://science.nasa.gov/media/medialibrary/2015/07/08/NASA-AAPR2015-FINAL.pdf>.
  58. Monya Baker, “Databases Fight Funding Cuts,” *Nature* 489 (2012): 19, doi:10.1038/489019a.
  59. Minoru Kanehisa, “Plea to Support KEGG,” *Kyoto Encyclopedia of Genes and Genomes*, 2011, <http://web.archive.org/web/20151108212102/http://www.genome.jp/kegg/docs/plea.html>.
  60. *Ibid.*
  61. National Institutes of Health, *Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research* (Bethesda, MD: National Institutes of Health, February 2015), <https://web.archive.org/web/20150908072046/https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf>.
  62. National Science Foundation, *Today’s Data, Tomorrow’s Discoveries*, NSF 15-52 (Arlington, VA: National Science Foundation, 2015), <https://web.archive.org/web/20160131120745/http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf>.
  63. Holdren, “Increasing Access to the Results of Federally Funded Scientific Research.”
  64. Borgman, “The Conundrum of Sharing Research Data.”
  65. *Ibid.*

66. Stewardship Gap Project, “The Problem,” 2015, [http://web.archive.org/web/20160214022939/http://www.colorado.edu/ibs/cupc/stewardship\\_gap/](http://web.archive.org/web/20160214022939/http://www.colorado.edu/ibs/cupc/stewardship_gap/).
67. David L. Vaux, “Research Methods: Know When Your Numbers Are Significant,” *Nature* 492 (2012): 180–81, doi:10.1038/492180a.
68. David Crotty, “Nevermind the Data, Where Are the Protocols?” *The Scholarly Kitchen* (blog), November 18, 2014, <https://web.archive.org/web/20160201153044/http://scholarlykitchen.sspnet.org/2014/11/18/nevermind-the-data-where-are-the-protocols/>.
69. National Institutes of Health, “Enhancing Reproducibility through Rigor and Transparency,” NOT-OD-15-103, October 30, 2015, <http://web.archive.org/web/20151030024011/http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-103.html>.

## Bibliography

- American Physical Society. “Supplemental Data.” *APS Online Style Manual*, 2003. <http://web.archive.org/web/20160325000016/http://www.apsstylemanual.org/oldmanual/parts/supplemental.htm>.
- Anderson, Nicholas R., Peter Tarczy-Hornoch, and Roger E. Bumgarner. “On the Persistence of Supplementary Resources in Biomedical Publications.” *BMC Bioinformatics* 7 (2006): 260. doi:10.1186/1471-2105-7-260.
- Baker, Monya. “Databases Fight Funding Cuts.” *Nature* 489 (2012): 19. doi:10.1038/489019a.
- Beebe, Linda. “Supplemental Materials for Journal Articles: NISO/NFAIS Joint Working Group.” *Information Standards Quarterly* 22, no. 3 (Summer 2010): 33–37. [http://web.archive.org/web/20151227001104/http://www.niso.org/apps/group\\_public/download.php/4885/Beebe\\_SuppMatls\\_WG\\_ISQ\\_v22no3.pdf](http://web.archive.org/web/20151227001104/http://www.niso.org/apps/group_public/download.php/4885/Beebe_SuppMatls_WG_ISQ_v22no3.pdf).
- Berg, Jeremy. “Indirect Cost Rate Survey.” *Data Hound* (blog), May 10, 2014. <https://web.archive.org/web/20150924200621/http://datahound.scientopia.org/2014/05/10/indirect-cost-rate-survey/>.
- Borgman, Christine L. “The Conundrum of Sharing Research Data.” *Journal of the American Society for Information Science and Technology* 63, no. 6 (June 2012): 1059–78. doi:10.1002/asi.22634.
- Borowski, Christine. “Enough Is Enough.” *Journal of Experimental Medicine* 208, no. 7 (2011): 1337. doi:10.1084/jem.20111061.
- Brown, C. Titus. “Cultural Confusions about Data: The Intertidal Zone between Two Styles of Biology.” *Living in an Ivory Basement* (blog), April 2, 2015. <http://web.archive.org/web/20151003163047/http://ivory.idyll.org/blog/2015-culturally-confused-about-data.html>.
- Burroughs Wellcome Fund. “Obtaining Tenure.” Career Tools, 2014. <http://web.archive.org/web/20151024181949/http://www.bwfund.org/career-tools/obtaining-tenure>.
- Burwell, Sylvia M., Steven VanRoekel, Todd Park, and Dominic J. Mancini. “Open Data Policy: Managing Information as an Asset.” Memorandum for the Heads of Executive Departments and Agencies, Office of Management and Budget, May 9, 2013. <https://web.archive.org/web/20151104005245/https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.

- Carpenter, Todd. "Supplementary Materials: A Pandora's Box of Issues Needing Best Practices." *Against the Grain* 21 (2009): 84–85.
- CPANDA. "What Is CPANDA?" 2015. <http://web.archive.org/web/20151120151725/http://www.cpanda.org/cpanda/about>.
- Crotty, David. "Nevermind the Data, Where Are the Protocols?" *The Scholarly Kitchen* (blog), November 18, 2014. <https://web.archive.org/web/20160201153044/http://scholarly-kitchen.sspnet.org/2014/11/18/nevermind-the-data-where-are-the-protocols/>.
- Decker, Robert S., Leslie Wimsatt, Andrea G. Trice, and Joseph A. Konstan. *A Profile of Federal-Grant Administrative Burden among Federal Demonstration Partnership Faculty*. A Report of the Faculty Standing Committee of the Federal Demonstration Partnership. Federal Demonstration Partnership, January 2007. [http://web.archive.org/web/20160214195603/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga\\_054586.pdf](http://web.archive.org/web/20160214195603/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_054586.pdf).
- Ember, Carol, and Robert Hanisch. "Sustaining Domain Repositories for Digital Data: A White Paper," 2013. [http://datacommunity.icpsr.umich.edu/sites/default/files/White-Paper\\_ICPSR\\_SDRDD\\_121113.pdf](http://datacommunity.icpsr.umich.edu/sites/default/files/White-Paper_ICPSR_SDRDD_121113.pdf).
- Ferguson, Liz. "How and Why Researchers Share Data (and Why They Don't)." *Exchanges* (blog), November 3, 2014. <https://web.archive.org/web/20160116150325/http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/>.
- Future of Structural Biology Committees. *Recommendations for Continued Investment in Structural Biology Following the Sunset of the Protein Structure Initiative*. Bethesda, MD: National Institute of General Medical Sciences, December 2014. <http://web.archive.org/web/20151013183712/http://www.nigms.nih.gov/News/reports/Documents/NIGMS-FSBC-report2014.pdf>.
- Galperin, Michael Y., Daniel J. Rigden, and Xosé M. Fernández-Suárez. "The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection." *Nucleic Acids Research* 43, no. D1 (2015): D1–5. doi:10.1093/nar/gku1241.
- Gold, Anna K. "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries." *D-Lib Magazine* 13, no. 9/10 (2007). <http://works.bepress.com/agold01/4/>.
- Herold, Philip. "Data Sharing among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications." *Journal of Librarianship and Scholarly Communication* 3, no. 2 (2015): eP1244. doi:10.7710/2162-3309.1244.
- Holdren, John P. "Increasing Access to the Results of Federally Funded Scientific Research." Memorandum for the Heads of Executive Departments and Agencies, Office of Science and Technology Policy, Executive Office of the President, February 22, 2013. [http://web.archive.org/web/20160115125401/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](http://web.archive.org/web/20160115125401/https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf).
- Inter-university Consortium for Political and Social Research (ICPSR). "Phase 5: Preparing Data for Sharing." In *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle*, 5th ed, 36–39. Ann Arbor, MI: ICPSR, 2012.
- IPAMM Working Group. *Impact of Proposal and Award Management Mechanisms: Final Report*. Arlington, VA: National Science Foundation, 2007. <http://web.archive.org/web/20151024203046/http://www.nsf.gov/pubs/2007/nsf0745/nsf0745.pdf>.

- Kanehisa, Minoru. "Plea to Support KEGG." *Kyoto Encyclopedia of Genes and Genomes*, May 16, 2011. <http://web.archive.org/web/20151108212102/http://www.genome.jp/kegg/docs/plea.html>.
- Kent, Stephen, Chryssa Kouveliotou, David Meyer, Richard H. Miller, David Schade, James Schombert, Alexander Szalay, and Suresh Santhana Vannan. *Report of the Astrophysics Archives Program Review for the Astrophysics Division, Science Mission Directorate*. NASA, May 6–8, 2015. <http://web.archive.org/web/20151107185033/http://science.nasa.gov/media/medialibrary/2015/07/08/NASA-AAPR2015-FINAL.pdf>.
- Kenyon, Jeremy, and Nancy R. Sprague. "Trends in the Use of Supplementary Materials in Environmental Science Journals." *Issues in Science and Technology Librarianship*, Winter 2014. doi:10.5062/F40Z717Z.
- Kim, Youngseek, and Melissa Adler. "Social Scientists' Data Sharing Behaviors: Investigating the Roles of Individual Motivations, Institutional Pressures, and Data Repositories." *International Journal of Information Management* 35, no. 4 (August 2015): 408–18. doi:10.1016/j.ijinfomgt.2015.04.007.
- Marcus, Emilie. "Taming Supplemental Material." *Neuron* 64, no. 1 (October 2009): 3. doi:10.1016/j.neuron.2009.09.046.
- Maunsell, John. "Announcement Regarding Supplemental Material." *Journal of Neuroscience* 30 (2010): 10,599–600.
- Mischo, William, Mary Schlembach, and Megan O'Donnell. "An Analysis of Data Management Plans in University of Illinois National Science Foundation Grant Proposals." *Journal of eScience Librarianship* 3, no. 1 (2014). doi:10.7191/jeslib.2014.1060.
- Mons, Barend. "Open Science as a Social Machine: Where (the...) Are the Data?" [keynote address, International Digital Curation Conference, Amsterdam, the Netherlands, February 22–25, 2016], <http://www.dcc.ac.uk/sites/default/files/documents/IDCC16/Keynotes/Barend%20Mons.pdf>.
- Myers, Robert W., and Robert H. Abeles. "Conversion of 5-S-Methyl-5-Thio-D-Ribose to Methionine in *Klebsiella Pneumoniae*. Stable Isotope Incorporation Studies of the Terminal Enzymatic Reactions in the Pathway." *Journal of Biological Chemistry* 265 (1990): 16,913–21.
- National Institutes of Health. *Data Sharing Workbook*. Bethesda, MD: National Institutes of Health, last revised February 13, 2004. [http://web.archive.org/web/20151116180118/https://grants.nih.gov/grants/policy/data\\_sharing/data\\_sharing\\_workbook.pdf](http://web.archive.org/web/20151116180118/https://grants.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf).
- . "Definitions of Criteria and Considerations for Research Project Grant (RPG/X01/R01/R03/R21/R33/R34) Critiques." National Institutes of Health Grants and Funding, last updated March 21, 2016. [http://web.archive.org/web/20160325020441/https://grants.nih.gov/grants/peer/critiques/rpg\\_D.htm](http://web.archive.org/web/20160325020441/https://grants.nih.gov/grants/peer/critiques/rpg_D.htm).
- . "Enhancing Reproducibility through Rigor and Transparency," NOT-OD-15-103. October 30, 2015. <http://web.archive.org/web/20151030024011/http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-103.html>.
- . *Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research*. Bethesda, MD: National Institutes of Health, February 2015. <https://web.archive.org/web/20150908072046/https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf>.
- . "Success Rates and Funding Rates: Research Project Grants: Competing Applications, Awards, and Success Rates 1998–2014." NIH Data Book, 2015. <http://web>.

- archive.org/web/20151022204459/http://report.nih.gov/NIHDatabook/Charts/Default.aspx?showm=Y&chartId=124&catId=13.
- National Science Foundation. *Data Management for NSF EHR Directorate: Proposals and Awards*. Arlington, VA: National Science Foundation Education and Human Resources Directorate, March 2011. <http://web.archive.org/web/20151030234444/http://www.nsf.gov/bfa/dias/policy/dmpdocs/ehr.pdf>.
- . *Report to the National Science Board on the National Science Foundation's Merit Review Process: Fiscal Year 2013*. Arlington, VA: National Science Foundation, May 2014. <http://web.archive.org/web/20151022211742/https://www.nsf.gov/nsb/publications/2014/nsb1432.pdf>.
- . *Today's Data, Tomorrow's Discoveries: Increasing Access to the Results of Research Funded by the National Science Foundation*. NSF 15-52. Arlington, VA: National Science Foundation, 2015. <https://web.archive.org/web/20160131120745/http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf>.
- Pampel, Heinz, Paul Vierkant, Frank Scholze, Roland Bertelmann, Maxi Kindling, Jens Klump, Hans-Jürgen Goebelbecker, Jens Gundlach, Peter Schirmbacher, and Uwe Dierolf. "Making Research Data Repositories Visible: The re3data.org Registry." *PLoS ONE* 8, no. 11 (2013): e78080. doi:10.1371/journal.pone.0078080.
- Renear, Allen H., Simone Sacchi, and Karen M. Wickett. "Definitions of *Dataset* in the Scientific and Technical Literature." *Proceedings of the American Society for Information Science and Technology* 47, no. 1 (2010): 1–4. doi:10.1002/meet.14504701240.
- Sapp, Curtis, Michael Lord, and E. Roy Hammarlund. "Sodium Chloride Equivalents, Cryoscopic Properties, and Hemolytic Effects of Certain Medicinals in Aqueous Solution III: Supplemental Values." *Journal of Pharmaceutical Sciences* 64, no. 11 (November 1975): 1884–86. doi:10.1002/jps.2600641132.
- Savage, Caroline J., and Andrew J. Vickers. "Empirical Study of Data Sharing by Authors Publishing in PLoS Journals." *PLoS ONE* 4, no. 9 (2009): e7078. doi:10.1371/journal.pone.0007078.
- Schaffer, Thomas, and Kathy M. Jackson. "The Use of Online Supplementary Material in High-Impact Scientific Journals." *Science and Technology Libraries* 25, no. 1–2 (2004): 73–85. doi:10.1300/J122v25n01\_06.
- Schneider, Sandra L., Kirsten K. Ness, Sara Rockwell, Kelly Shaver, and Randy Brutkiewicz. *2012 Faculty Workload Survey*. Research report. Federal Demonstration Partnership, April 2014. [http://web.archive.org/web/20151022202705/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pgasite\\_087667.pdf](http://web.archive.org/web/20151022202705/http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pgasite_087667.pdf).
- Schwarzman, Alexander. "Supplemental Information: Who's Doing What and Why." PowerPoint presented at the CSE 2012 Annual Meeting, Seattle, WA, May 20, 2012. [http://web.archive.org/web/20151013195224/http://www.niso.org/apps/group\\_public/document.php?document\\_id=8591&wg\\_abbrev=supptechnical](http://web.archive.org/web/20151013195224/http://www.niso.org/apps/group_public/document.php?document_id=8591&wg_abbrev=supptechnical).
- . "Supplemental Materials Survey." *Information Standards Quarterly* 22, no. 3 (Summer 2010): 23–26. [http://web.archive.org/web/20151013194617/http://www.niso.org/apps/group\\_public/download.php/4886/Schwarzman\\_SuppMatlsSurvey\\_ISQ\\_v22no3.pdf](http://web.archive.org/web/20151013194617/http://www.niso.org/apps/group_public/download.php/4886/Schwarzman_SuppMatlsSurvey_ISQ_v22no3.pdf).
- Siebert, Sabina, Laura M. Machesky, and Robert H. Insall. "Overflow in Science and Its Implications for Trust." *eLife* 4 (2015): e10825. doi:10.7554/eLife.10825.
- Singh, Ajai R., and Shakuntala A Singh. "Ethical Obligation towards Research Subjects." *Mens Sana Monographs* 5, no. 1 (2007): 107–12. doi:10.4103/0973-1229.32153.

- Stewardship Gap Project. "The Problem." 2015. [http://web.archive.org/web/20160214022939/http://www.colorado.edu/ibs/cupc/stewardship\\_gap/](http://web.archive.org/web/20160214022939/http://www.colorado.edu/ibs/cupc/stewardship_gap/).
- Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. "Data Sharing by Scientists: Practices and Perceptions." *PLoS ONE* 6, no. 6 (2011): e21101. doi:10.1371/journal.pone.0021101.
- Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." *PLoS ONE* 10, no. 8 (2015): e0134826. doi:10.1371/journal.pone.0134826.
- US Geological Survey. "Scientific Data Management Foundation." U.S. Geological Survey Instructional Memorandum No. IM OSQI 2015-01. US Geological Survey Manual, February 19, 2015. <http://web.archive.org/web/20151031000659/http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-01.html>.
- Vale, Ronald D. "Accelerating Scientific Publication in Biology." Preprint. bioRxiv, September 12, 2015. doi:10.1101/022368.
- . "Accelerating Scientific Publication in Biology." *Proceedings of the National Academy of Sciences* 112 (2015): 13,439–46. doi:10.1073/pnas.1511912112.
- Vaux, David L. "Research Methods: Know When Your Numbers Are Significant." *Nature* 492 (2012): 180–81. doi:10.1038/492180a.
- Wald, Chelsea. "Redefining Tenure at Medical Schools." *Science Careers*, March 6, 2009. doi:10.1126/science.caredit.a0900032.
- Wicherts, Jelte M., Denny Borsboom, Judith Kats, and Dylan Molenaar. "The Poor Availability of Psychological Research Data for Reanalysis." *American Psychologist* 61, no. 7 (October 2006): 726–28. doi:10.1037/0003-066X.61.7.726.
- Williams, Sarah C. "Data Sharing Interviews with Crop Sciences Faculty: Why They Share Data and How the Library Can Help." *Issues in Science and Technology Librarianship*, Spring 2013. doi:10.5062/F4T151M8.