

## Supporting the Proliferation of Data-Sharing Scholars in the Research Ecosystem

Abstract: Librarians champion the value of openness in scholarship and have been powerful advocates for the sharing of research data. College and university administrators have recently joined in the push for data sharing due to funding mandates. However, the researchers who create and control the data usually determine whether and how data is shared, so it is worthwhile to look at what they are incentivized to do. The current scholarly publishing landscape in addition to the promotion and tenure process creates a “prisoner’s dilemma” for researchers as they decide whether to share data. This is consistent with the observation that researchers in general are eager for others to share data, but reluctant to do so themselves. If librarians encourage researchers to share data and promote openness without simultaneously addressing the academic incentive structure, those who are intrinsically motivated to share data will be selected against via the promotion and tenure process. This will cause those who do not share to be disproportionately recruited into the senior ranks of academia. To mitigate the risk of this unintended consequence, librarians must advocate for a change in incentives alongside the call for greater openness. Highly cited datasets must be given similar weight to highly cited articles in promotion and tenure decisions in order for researchers to reap the rewards of their sharing. Librarians can help by facilitating data citation to track the impact of datasets and working to persuade administrators of the value of rewarding data sharing in tenure and promotion.

## Supporting the Proliferation of Data-Sharing Scholars in the Research Ecosystem

Librarianship is as much about values as it is about skills and experience. In the conversation surrounding how research data is to be managed, preserved, and shared, it is no surprise that librarians have been strong advocates of openness. Right now, we find ourselves in a rare moment where many of the institutions within which we work are pushing in the same direction. Recognizing the tremendous investment the public has made in research data and the need for accountability, funding agencies have insisted that data be better preserved and shared more openly. Colleges and universities, wanting their researchers to remain competitive for funding opportunities, have begun to shift their policies towards data accordingly; some have vested librarians with substantial responsibility for providing data services. As a general rule, people respond to changes in incentives, and institutions, large and slow moving as they may be, do likewise.

My purpose here is to urge librarians to keep this in mind as we work to shape data policy and support researchers as they try to navigate it. Guided as we are, and rightly so, by our values, we need to be mindful not to create perverse incentives in the name of openness that end up sabotaging genuine progress toward it. If we want accessible data and the benefits it brings to the research enterprise, we need to take a step back and examine the process that gets us there. Specifically, we need to look at the researchers themselves, who for the most part individually determine whether and how data is shared. What defines the landscape of their practice?

Generally speaking, academic researchers operate within an ecosystem comprised of other such researchers who decide whether they will succeed or fail within their fields. Success is defined here as a long and productive career, which is made possible through the promotion and tenure process. Not all participants in the ecosystem get to advance – as in a natural system, resources are limited, so those who produce what their peers value become the “winners” while others are left behind to eventually drop out of the field. For better or for worse, publications in highly cited journals are still overwhelmingly valued over other research outputs, and the more, the better.

We are then faced with the empirical question of the relationship between sharing data and publishing papers. A mathematical model of incentives in data sharing (Pronk et al. 2015) suggests that from the individual perspective, the decision to share or not constitutes a prisoner’s dilemma. Briefly, a prisoner’s dilemma is a situation where two or more actors must choose between cooperating with each other and defecting (refusing to cooperate, betraying the

partner). Cooperation produces the best overall outcome, measured in terms of the benefit to all actors combined, but defecting produces the optimal outcome for each individual because of the steep penalty for cooperating when others defect, known as the sucker's payoff. This means that defection emerges as the winning strategy, reducing the total systemic benefit. To put this in terms of openness, those who make their data available are cooperating, whereas those who do not make their data available are defecting. The more people who cooperate, the more productive the ecosystem—data sharing allows advances to happen faster and results in more publications communicating those discoveries. However, those who defect and use the data others share without sharing their own always do better individually, outpacing their competitors in number of papers published.

A key insight from Pronk et al.'s (2015) simulation is that the non-sharers always outperformed those who shared regardless of changes in the model's parameters. This remained the case even when the cost to share data was reduced (if, for instance, sharing took less of researchers' time and effort) or, crucially, even when the overall level of sharing in the system rose – that is to say, whether nearly everyone or nearly no one shared did not change the fact that non-sharers, individually, had a competitive advantage over sharers. I highlight this result because it implies that to the extent the model accurately reflects the real ecosystem, appealing to researchers' values and community spirit to encourage them to share (Darch and Knox 2017) will not be effective if our goal is to foster long-term change in scholarly norms surrounding openness. In fact, it could be counterproductive.

To extend the biological metaphor, consider the researchers within a field as an evolving population, and suppose that some individuals are more inclined towards sharing data than others. Let us also assume that senior researchers are in a better position to enforce scholarly norms and values through institutional power. Since not all junior researchers receive tenure and promotion, selection pressures (operating similarly to natural selection) can develop within the system for or against various traits that affect the likelihood of advancement. As the system currently stands, since those who defect outperform those who cooperate, there is selection pressure against those who are committed to sharing as a matter of principle. Consequently, we would expect to see fewer such individuals as time passes and the cohort advances to higher ranks. This is the opposite of what we would prefer and leaves little hope that future leaders in a field will contribute to the normalization of open data practices. If librarians encourage individual researchers to change their behaviors without simultaneously addressing the incentives that define the ecosystem, we are contributing to the selection pressure against those who are internally motivated to do the right thing.

The strongest evidence that the model described above reflects reality is the observation that while junior researchers express the most support for data sharing and reuse as a community standard, they are less likely to share their own data than senior researchers (Tenopir et al. 2015). This does not mean that these researchers are consciously being mercenary. In fact, some are frustrated that they cannot be “good citizens” in the scientific community without hurting their careers. In a recent Twitter discussion, a cognitive neuroscientist solicited advice on behalf of early career researchers supportive of open science who were concerned it hampered their productivity relative to competitors who repeatedly published less transparent studies in “top” journals (Chambers 2017). As advocates for openness, we recognize that in order for institutional change to last, future leaders must be at least as committed as current ones, if not more. But expecting those positioned most precariously within the system to cooperate regardless of personal consequences will not achieve our goals. Further, I believe that it is unjust.

How can we change the system so that data sharing is professionally rewarded? First, we need to get better at communicating its value in terms of the coin of the research realm, citations. Although academics within and outside the library alike view quantitative measures of performance with suspicion since they can be a source of perverse incentives, metrics are often necessary to get the attention of administrators. The possibility of increased citations for open-access articles has helped to balance out the added burden of OA mandates for publications. Something similar needs to happen for data, and that depends on our ability to track citations of datasets as we currently do articles. Demonstrating the impact of data on scholarship quantitatively is necessary (though not sufficient) to making the case that shared data should receive consideration in promotion and tenure decisions.

With our supporting evidence in place, I believe we should take advantage of the current alignment between librarians’ push for openness and a renewed institutional focus on compliance with new mandates in this area. While academic departments may be loath to relinquish any control of their internal promotion processes, higher administrators are in an ideal position to ask the hard questions about why the products of research they want to reward are not being counted fairly in tenure, promotion, and hiring. Librarians have been at the vanguard of those demanding greater faculty accountability for poor outcomes in diversity, equity, and inclusion. As unappealing as more administrative oversight of academic promotion might be, we can argue that it is justified due to the lack of progress in those areas. This principle can likewise be extended to the issue of data sharing in particular and openness in general. Librarians are at once part of the academy and uniquely positioned within it to see the bigger

picture. What I have personally observed is a research ecosystem that will not change so long as it remains largely closed, self-contained, and self-reinforcing. Now is the time for librarians to be forceful advocates for openness in conversation and cooperation with those institutional stakeholders best positioned to perturb the system from the outside. Only then will researchers who habitually share data advance through their fields in sufficient numbers to ensure lasting change.

## References

Chambers, Chris. Twitter Post. April 24, 2018, 1:01 PM.

<https://twitter.com/chrisdc77/status/988840333625778176>.

Darch, Peter T., and Emily J. M. Knox. 2017. "Ethical Perspectives on Data and Software Sharing in the Sciences: A Research Agenda." *Library & Information Science Research* 39 (4): 295–302. <https://doi.org/10.1016/j.lisr.2017.11.008>.

Pronk, Tessa E., Paulien H. Wiersma, Anne van Weerden, and Feike Schieving. 2015. "A Game Theoretic Analysis of Research Data Sharing." *PeerJ* 3 (September): e1242. <https://doi.org/10.7717/peerj.1242>.

Tenopir, Carol, Elizabeth D. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." *PLOS ONE* 10 (8): e0134826. <https://doi.org/10.1371/journal.pone.0134826>.