American Psychological Association

American Psychologist

In the public domain

2018, Vol. 73, No. 2, 126–137
http://dx.doi.org/10.1037/amp0000197

Can Psychology Walk the Walk of Open Science?

Bradford W. Hesse
National Cancer Institute, National Institutes of Health, Bethesda, Maryland

An "open science movement" is gaining traction across many disciplines within the research enterprise but is also precipitating consternation among those who worry that too much disruption may be hampering professional productivity. Despite this disruption, proponents of open data collaboration have argued that some of the biggest problems of the 21st century need to be solved with the help of many people and that data sharing will be the necessary engine to make that happen. In the United States, a national strategic plan for data sharing encouraged the federally funded scientific agencies to (a) publish open data for community use in discoverable, machine-readable, and useful ways; (b) work with public and civil society organizations to set priorities for data to be shared; (c) support innovation and feedback on open data solutions; and (d) continue efforts to release and enhance high-priority data sets funded by taxpayer dollars. One of the more visible open data projects in the psychological sciences is the presidentially announced "Brain Research Through Advancing Innovative Neurotechnologies" (BRAIN) initiative. Lessons learned from initiatives such as these are instructive both from the perspective of open science within psychology and from the perspective of understanding the psychology of open science. Recommendations for creating better pathways to "walk the walk" in open science include (a) nurturing innovation and agile learning, (b) thinking outside the paradigm, (c) creating simplicity from complexity, and (d) participating in continuous learning evidence platforms.

Keywords: data sharing, open science, big data, culture of science

All of the life sciences, psychology included, are struggling with the implications of an "open science" movement, which feels "destructive" of the status quo for some (Topol, 2013) but feels like the dawning of a golden age of "information liberation" for others (Park, 2011, p. 45). By "open science" I refer to the social and epistemological movement—enabled by the tools of the information age—to make the publication of scientific concepts together with the protocols and data upon which those concepts are based readily accessible to all levels of an inquiring society. In a poignant example of how pervasive the open science discussion has become, at least in the biomedical sciences, then-Vice President of the United States Joe Biden addressed the topic directly in his comments to a conference of 138,000 oncologists and medical researchers on June 6, 2016 (Ong, 2016). The speech was a follow-up on the president's publicly announced "Cancer Moonshot Initiative," so-named because the White House wanted to pull out all the stops in doubling the nation's pace of progress against the national burden of cancer. Within the biomedical community, there has been a growing recognition that the only way to crack the code on cancer is to share data as broadly as possible, with the hope that a collective analytic view could reveal patterns inaccessible to isolated research laboratories. In speaking candidly with the gathered audience of practitioners and professionals, the vice president summed up this dawning awareness as follows:

> To be honest with you, it [i.e., accelerating the pace of cancer research and oncology practice] . . . requires a change in mindset. It requires a lot more openness: open data, open collaboration, and above all, open minds. (pp. 1-8)

This challenge to the research and practice communities was tinged with no small degree of frustration in realizing that a vision of an open, collaborative research platform in medicine seems to have taken so long to achieve. Yet at the

Editor's note. This article is part of a collection titled "Data Sharing in Psychology," published as a special section of *American Psychologist* (February-March, 2018). Jennifer Crocker and Leah Light served as editors of the special section.

Work on this article was conducted as part of the author's official duty as branch chief for the Health Communication and Informatics Research Branch within the Behavioral Research Program at the National Cancer Institute. Nevertheless, opinions presented within the article should be attributed solely to the author and do not necessarily represent those of the National Cancer Institute or the National Institutes of Health.

Correspondence concerning this article should be addressed to Bradford W. Hesse, Division of Cancer Control and Population Sciences, Behavioral Research Program, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Room 3E610 MSC 9761, Bethesda, MD 20892-7365. E-mail: hesseb@mail.nih.gov

Bradford W. Hesse

same time, the speech was filled with hope that biomedical science had reached a tipping point in removing barriers to data sharing and interdisciplinary collaboration, a tipping point that could move science more quickly toward finding life-saving cures and treatments. That hope was further instantiated through a robustly bipartisan passage of the 21st Century Cures Act in December 2016 (21st Century Cures Act, 2015), which set into play congressionally authorized incentives for data sharing and penalties for data blocking (Hudson & Collins, 2017).

In this article, I trace some of the origins of the open science culture as embraced by the vice president's comments, but I then take a hard, candid look at how feasible it may be to adhere to the precepts of a fully open scientific culture in psychology. Undoubtedly, there are certain very real obstacles that must be overcome before the field is ready to adopt a massive culture change in the ways in which it shares and publishes the products of its constituent scientists. Nevertheless, there are examples of data-sharing platforms that have been constructed and are thriving in the field. These examples should provide worthy fodder for evaluating the incentives for and against data sharing in the social sciences. I then end the article with recommendations for moving forward, perhaps not in every aspect of a collective science in psychology but at least in some areas where the stakes for inclusion are the highest.

The Open Science Movement

In many respects, the idea of open science is not new but can be traced back to the massive culture shift occurring in the 15th century when Johannes Gutenberg introduced a way of mass-producing written text using the printing press and movable type. Many sociologists of science credit the Gutenberg invention, which served as a significant disruption to the status quo, for helping to usher in a renaissance in scholarly thinking and for setting the foundation upon which the modern scientific enterprise could be built. It was, as Marshal McLuhan would later observe, a prototypical example of how new communication technologies—the printing press, radio, telephone, movies, TV, and now the Internet—would have a profound influence on the ways in which individuals and cultures cognitively processed and organized information from the world around them (McLuhan, Gordon, Lamberti, & Scheffel-Dunand, 2011). In the 21st century, the Internet has risen as the new paradigm-shifting technology for science, built on a model of distributed support for an invisible college (Crane, 1972) of investigators working in parallel to each from around the globe to answer pressing scientific questions (Hesse, Sproull, Kiesler, & Walsh, 1993; Sproull & Kiesler, 1991).

The integration of this new infrastructure into the fabric of science has rekindled a spate of conversations about what is possible, what is advisable, and what is desirable in the zeitgeist of the prevailing scientific culture. These conversations often talk past each other, because their arguments stem from different frames of reference or are oriented toward solving different aspects of the same problem. Fecher and Friesike reasoned that there were predominantly five different schools of thought governing these conversations (Fecher & Friesike, 2014). I review each briefly as a prelude for considering the applicability of open science principles in the field of psychology.

The infrastructure school of thought. Some of the discussions in the published literature represent a call for action to develop the architectures needed to support open and collaborative processes in research and practice. These are infrastructure-oriented discussions, focused on understanding what the engineering requirements should be for platforms designed to improve scientific collaboration and data sharing. Douglas Engelbart was one of the first progenitors of this school of thought, because he outlined a blueprint in the early 1960s for how computer architectures could "augment human intellect" and improve open collaboration among teams of scientists (Engelbart, 1961, p. 3). This early blueprint led to some of the initial prototypes for what is now the Internet, used primarily among agencies funded by the Department of Defense. In the late 1980s a series of papers and reports commissioned by the National Science Foundation (NSF) assessed the potential contributions of infrastructure for science (e.g., National Academy of Sciences, 1989). The reports' persuasive arguments helped lead the NSF to invest in a national backbone for electronic data interchange between its funded supercomputer centers—a backbone that would eventually be privatized and that would lead to the rapid diffusion of the World Wide Web in

the years to follow. In 1990, the National Institutes of Health (NIH) funded the Human Genome Project to determine the full sequence of nucleotide base pairs that constitute human DNA. Substantial investments were made into the networked data repository, GenBank, to hold the 3 billion-plus base pairs contributed by thousands of laboratories in parallel from around the world. In 2001, the National Committee on Vital and Health Statistics laid out a vision for developing the National Health Information Infrastructure to connect the nation's hospitals, public health surveillance systems, and research laboratories (National Committee on Vital and Health Statistics, 2001). Infrastructure discussions of this type have helped set the agenda for continued investment in academic computing centers, high-throughput data lines, electronic health records (EHRs), advanced data-analytic tools, and subsidized access to Internet capacity for underresourced communities.

The public good school of thought. In contrast to the technological focus of infrastructure discussions, arguments pitched from a public good school of thought tend to emphasize the societal value of opening access to the products of science for those in the public sphere who can take best advantage of them. Congress, as a representative of the public interest and the appropriations authority for scientific agencies, has begun to weigh in more broadly on discussions surrounding open access to the products of science funded by taxpayer dollars. As an illustrative case in point, the Fair Access to Science and Technology Research Act was introduced to the Senate in 2013 (Fair Access to Science and Technology Research Act, 2013) and again in 2015 (Fair Access to Science and Technology Research Act, 2015). It stipulated that publicly funded research publications should be made freely available to the public a short time (interpreted by the NIH currently as 12 months) after its publication (National Institutes of Health, 2016). Shortly after the bill was introduced, the Executive Branch's Office of Science and Technology Policy issued a memorandum directing federal agencies with research expenditures over $100 million to develop plans for supporting public access to the products of federally funded research (Holdren, 2013). The memorandum directed the agencies to (a) publish open data in a discoverable, machine-readable, useful way; (b) work with the public and civil society organizations (such as the American Psychological Association [APA]) to prioritize open data sets for release; (c) support innovation to improve open data based on feedback; and (d) continue to release and enhance high-priority data sets, especially through the open access web portal data.gov.

The democratic school of thought. Like discussions around public good, scientific associations and publishers routinely engage in conversations about open science from the perspective of providing equitable access to members of their respective scientific communities. Normal science, in this vein, is viewed as an accretive enterprise in which individual members contribute ideas and findings openly for peer review and collective evaluation. Withholding scientific information from the community weakens the collective enterprise by preventing the community from engaging in the necessary replication or transparent testing needed to advance the field. Paper-based publishing may have inadvertently placed limits on full involvement by the community, effectively biasing the scientific record by forcing null findings into the file drawer, restricting access to underlying data, and limiting dissemination to those institutions with the financial resources to afford premium journal access (Dalton, Aguinis, Dalton, Bosco, & Pierce, 2012; Ioannidis, 2016; Rosenthal, 1979). Proponents of open science in this regard have encouraged open access, intellectual property rights, open data, and open code as the means for improving democratic access to knowledge for members within their respective guilds (e.g., Nosek et al., 2015).

The pragmatic school of thought. Closely allied to arguments from the democratic school is the practical observation that science can simply be made more efficient if the community were to engage in a more effectual use of scarce resources. Consider the case of Wikipedia, a crowdsourcing technology platform that within 16 years has built up an archive of 5,326,744 articles, delivered in 295 languages, and with a level of scholarly accuracy matching that of the Encyclopedia Britannica in a head-to-head comparison (cf. Giles, 2005). The platform works so efficiently because it takes direct advantage of massively distributed, parallel input from a broad community of contributors. The same is beginning to happen in science, as distributed communities of researchers synchronize their work along parallel paths to accelerate progress within their disciplines (Nielsen, 2012). Open science discussions from the pragmatic school of thought focus on the ways in which cyberinfrastructure platforms can be used to accelerate understanding in such broad areas as brain mapping, molecular medicine, meteorology, physical oceanography, group behavior, and global economics to name just a few.

The measurement school of thought. If scientific work is going online, then there should be new ways of measuring scientific impact in an open science environment. Web analytics tools allow publishers to quantify the number of times an article may be viewed, opened, downloaded, or shared through social media. New services have begun to emerge for quantifying these online behaviors into alternative metrics of scientific productivity and to offer professional credit for donating and documenting data, for registering study parameters before conducing the research, and for conducting replication studies (Chavan & Penev, 2011; Gorgolewski, Margulies, & Milham, 2013). Technical specialists who curate content within the APA's own PsycINFO bibliographic information service have been experimenting with methods for providing users with tools to trace the thematic lineage of scientific findings through

cited references, of improving access to the scientific canon through expanding coverage, of disambiguating authors' attributions through ORCID, and of sorting articles by frequency of citation. All of these efforts—and others—are aimed at helping to elevate new conversations regarding the ways in which communities assess the contributions of their members in the age of the Internet (Cooper & VandenBos, 2013b).

The Science of Psychology and the Psychology of Science

Across these schools of thought, psychology has been playing an understated but crucial role in two ways. First, many participants in the funded community of psychological research have already been experimenting with at least some aspects of an open science culture. It is instructive to learn as much as possible from these early experiments to understand how compatible at least some aspects of the discipline might be with open science principles. Second, because science itself is a human endeavor, psychological evidence can contribute to discussions on how to design the tools and artifacts of scientific research to overcome issues of misaligned incentives, of collaborative dysfunction, of friction in sharing data and even to promotion of a sense of data altruism among participants. Interweaving discussions from each of the schools of thought from the open science movement with the evidence being created by psychological scientists and practitioners will enrich the discourse across disciplines.

Open Science in Psychology

Three years before the U.S. president announced a moonshot initiative for cancer, he made a similarly far-reaching commitment in 2013 to begin a project that would rival the human genome endeavor by committing funds "to map the circuits of the brain, measure the fluctuating patterns of electrical and chemical activity flowing within those circuits, and understand how their interplay creates our unique cognitive and behavioral capabilities" (Choudhury, Fishman, McGowan, & Juengst, 2014, p. 1). The endeavor would be called "Brain Research Through Advancing Innovative Neurotechnologies," or the BRAIN initiative, and it would place the field of neuroscience in the same milieu of data sharing and open science as other disciplines within the life sciences.

It did not take long before discussions within the BRAIN community began to echo sentiments from the same schools of thought as others in the emerging open science culture. Those with a *pragmatic* view in the community would point to the expensive cost of neuroimaging technologies and argue that for efficiency's sake the yields from those investments should be placed in a common repository of scientific knowledge from which all laboratories could benefit (Landis et al., 2016). This would be especially true for ultrahigh field magnetic resonance images of the brain allowing for resolution at submillimeter levels. Sharing those scans would make it possible for other laboratories to develop and apply the analytic and pattern-making, machine-learning techniques needed to extract knowledge from high-density data (Tardif et al., 2016). Furthermore, the vast amounts of data produced by many of these new technologies—some have referred to "yottabytes" (or 10^{24}) of neuroimaging data—would outstrip the capacity of any single lab to store or to analyze (Choudhury et al., 2014). A common *infrastructure* would be needed to support a virtual scientific collaboratory in neuroscience, thereby defraying the expenses of any one university while taking advantage of scarce analytic resources by sharing analytic tools and by offering data up for community analysis (Gorgolewski et al., 2015; Landis et al., 2016; Wang et al., 2016; Wood et al., 2014).

Those arguing from a *democratic framework* have suggested that in the other biological sciences norms have been changing to favor inclusion. Not making data available to the broader community for reanalysis or replication is increasingly perceived as an ethical infraction, with punitive actions against those who are noncompliant and rewards meted out for those who are (Chavan & Penev, 2011; Reid et al., 2016). Advocates have also taken a *public accountability* stance, suggesting that researchers have an obligation to the patients contributing data from their neuroimaging scans to ensure that their contributions are used to accelerate discovery in neuroscience (Poldrack & Gorgolewski, 2014) and that all ethical obligations are met to nurture trust between the patient and research communities (Critchley, Bruce, & Farrugia, 2013).

Even before the announcement of the BRAIN initiative, there were historical precedents for data sharing within the psychological and social science communities (Hedrick, 1988). Pointing to the success of the GenBank project designed to host data for the human genome project, Van Horn and Gazzaniga described their rationale in 2002 for creating a virtual archive for functional magnetic resonance imaging (fMRI) scans to support new discoveries in neuroscience well in advance of the president's public commitment (Van Horn & Gazzaniga, 2002). The Interuniversity Consortium for Political and Social Science Research (ICPSR) has similarly been hosting microdata and data subset extraction tools from national surveys, from university-sponsored research, and from independent social scientists through their web portal for close to five decades (Chung, Mullner, & Yang, 2002). The American Psychological Association formed a partnership with ICPSR to serve as the trusted repository for data being deposited along with open access articles when it launched the Archives of Scientific Psychology (Cooper & VandenBos, 2013a).

The Center for Open Science was founded as a nonprofit technology organization to make the data, code, and resources of psychological researchers easily available to other psychologists for the purposes of transparency, replication, and improved rigor. These platforms have been useful in germinating replication studies to extend the reliability and generalizability of scientific discourse (Asendorpf et al., 2013; Nosek et al., 2015) and—for better or worse—in disrupting the status quo (Bohannon, 2014). It is too early to know just how rapidly these new open science platforms may be diffusing throughout the general scientific community. The diffusion curves will likely vary by discipline and may be likely to trend upward across all disciplines as researchers become familiar with popular file-sharing utilities such as Google Drive, Dropbox, Microsoft's Office 365, and Apple's iCloud. From one analysis, the number of research articles citing the use of one specific type of code-sharing platform (GitHub, which is explained later) rose from less than .1% to 1% from 2012 to 2016 in the computer sciences but also rose from about .05% to .2% in mathematics and from near 0% to .2% in the neurosciences (Perkel, 2016).

In an exercise initiated by the Office of Behavioral and Social Science Research and supported by multiple organizations (the National Cancer Institute, the Substance Abuse and Mental Health Services Administration, the Centers for Disease Control and Prevention, the Robert Wood Johnson Foundation, and the Pritzer Family Fund), the Institute of Medicine (now the National Academy of Medicine) identified the types of measures and data to be included routinely as fields within the increasingly ubiquitous distribution of electronic health records, or EHRs (Institute of Medicine, 2015). Including behavioral and psychological measures as routine data-collection points within health-care systems would provide two benefits back to the psychological community. First, it would ensure that psychosocial considerations would be built into the clinical decision-making process, thus addressing the needs of the "whole patient" (Institute of Medicine, 2008, p. ix). Second, it would make psychological data from practice-based EHRs available to researchers, administrators, and practitioners within the framework of a "learning healthcare system," a concept used by the Institute of Medicine to describe the much-needed feedback loop between evidence-based practice and practice-based evidence (Institute of Medicine, 2013, p. ix).

The Psychology of Open Science

Aside from articulating the benefits of data sharing from a national priority perspective, the President's Council of Advisors on Science and Technology also called for more social science research to understand how open science infrastructures can be molded to augment human cognition and to nudge support for collaborative behavior (President's Council of Advisors on Science and Technology, 2010a). Terms such as *crowdsourcing* (Handler & Ferrer Conill, 2016), *social computing* (Chandra, Iyer, & Raman, 2015), and *technology-mediated social participation* (Shneiderman, 2011) have all entered the common lexicon to describe what is essentially an unprecedented new phenomenon facilitated by electronic collaboration. New psychological and sociological research is needed, the council argued, to understand how the affordances of this new environment can be optimized to improve the capacity for data sharing and open collaboration in the sciences and in society more generally (President's Council of Advisors on Science and Technology, 2010a). This is an era in which human factors research, behavioral economics, social psychology, team science, and organizational psychology can all be brought to bear on the goal of improving the scientific enterprise (Buttliere, 2014).

A corollary to understanding the motivations and affordances needed to support data sharing among scientists is the need to understand how incentives and protections can be instituted for encouraging data donation from research participants (Woolley et al., 2016). The Robert Wood Johnson Foundation has referred to this as an opportunity to create public incentives to "donate personal data for public good" (California Institute for Telecommunications and Information Technology, 2014, p. 13), especially in light of the number of people who now own wearable devices (e.g., Fitbit, the Apple watch) and have expressed a willingness to share their quantified-self measurements with others (Bietz et al., 2016; California Institute for Telecommunications and Information Technology, 2014). The NSF has invested similarly in promoting tools for "citizen science" (Preece, 2016, p. 585), an area of study aimed at inviting laypersons in the field to assist in scientific measurements and observations. When citizens help by serving essentially as a distributed network of data collectors, they can accelerate the pace of observations geometrically (Nielsen, 2012). Internal discussions at the NSF relevant to the topic have benefitted from discussions with grantees funded by the Directorate for Social, Behavioral, and Economic Sciences (SBE; Pirolli, Preece, & Shneiderman, 2010). An emerging interest in Health 2.0, in which members of the public are taking an active interest in participating directly in their own health care and the health of their communities, has created an opportunity to receive donated data from patients for the superordinate goal of refining scientific knowledge about the efficacy and effectiveness of treatments (Hesse et al., 2010).

Obstacles to Open Science in Psychology

Despite progress in adopting open science principles within areas such as neuroscience, many impediments remain before diffusing the practice within the social sciences more generally. One aspect is practicality. Reports have circulated within other disciplines of grand and ambitious efforts to create inclusive repositories for sharing data sets,

but once the skeletal repository was built and the call went out for volunteered data, almost none came in (Nelson, 2009). The practical reasons were plentiful. Researchers often had a difficult time going back over archived studies, finding the accompanying data sets, getting them ready for sharing, and then going through the operational process of uploading the data sets into suitable repositories. Even when older data sets were found, they were often in a format that was no longer compatible with contemporary data technologies. They might have been on reel-to-reel mainframe tapes, 5[1/4]-in. or 3[1/2]-in. floppy disks, Zip drives, or some other sort of obsolete medium. In many cases, the original data may have been encoded through obsolete software or unsupported data programs.

Even if the electronic media were readable, the data formats would still need to be documented sufficiently to allow others to utilize the information within a follow-on study. Security and confidentiality issues would have to be addressed with appropriate measures taken to deidentify the data where appropriate, to withhold (or suppress) statistics when counts are low and individuals or groups are at risk of being identified, and to assure that participants' trust in the confidentiality of their personal information is honored (Joly, Dalpe, So, & Birko, 2015; Pearce & Smith, 2011; Woolley et al., 2016). Institutional review boards, especially as distributed across the multiple institutions in a consortium, would need to find ways of streamlining their procedures and to find a suitable solution for secondary usage of data beyond the parameters foreseen in the original consent document (Piwowar, Becich, Bilofsky, & Crowley, 2008). Unfortunately, scientists are rarely given the resources to do all this work after the fact, which puts fulfilling data requests into the category of responding to an unfunded mandate. Without the resources or professional incentives to engage in data sharing, the task simply falls by the wayside in terms of priorities (Nelson, 2009; Pearce & Smith, 2011; Tenopir et al., 2011).

Another tenet of the open science movement that has been difficult to support in the behavioral research community is the goal of assembling individual contributions into a broader, thematically cohesive whole for knowledge aggregation. The human genome project from the biological sciences presents an example of how data can be integrated at scale. By carefully crafting a semantically cohesive lexicon of terms ahead of time, called an "ontology" in computer science terms, curators of the NIH-sponsored GenBank repository made it feasible to upload nucleotide sequences into a consistent format across laboratories (Benson et al., 2013). Such universal ontologies have generally been absent in the parochially divided communities of research in psychology (Bangdiwala et al., 2016; Larsen et al., 2017). Moving toward a more interoperable set of ontologies in behavioral research could help overcome the Balkanization of findings that repeatedly interferes with a translation from basic science to practice (Larsen et al., 2017; Michie et al., 2016). It can also help expand the frontiers of interdisciplinary science by allowing researchers to integrate behavioral data with other types of data for a broader view of phenomena in context (Bainter & Curran, 2015). More pragmatically, standards in terminology can help assure a place for mental health measures as an accompaniment to the biomedical measures being deposited within the data infrastructures funded by federal and state agencies under health-care-reform initiatives (Institute of Medicine, 2015).

A notable contribution in this arena was initiated by the National Institute of Mental Health (NIMH) in response to its 2008 strategic plan calling for a more precise way of classifying mental disorders based on dimensions of observable behavior and neurobiological measures (Cuthbert & Kozak, 2013). The goal of the program, referred to as the Research Domain Criteria (RDoC) project, was to enable a measurement system made up of shared, scientifically derived constructs spanning units of analysis from genes to molecules, cells, circuits, physiology, behavior, self-report, and paradigms. By utilizing constructs based on neurobiology and genomics as well as observable behavior, the system could then be integrated into the larger network of semantic relationships curated by the National Library of Medicine through its Medical Subject Headings (MeSH) and its Unified Medical Language System (UMLS). Doing so could help transform the behavioral and social sciences by moving researchers from their traditional footing within a data-lean environment onto a broader footing within an interconnected, data-rich environment (Collins & Riley, 2016).

Perhaps the most challenging obstacle to adopting open science principles in psychology, though, is simply the intransigence of tradition and the fear associated with breaking from tradition (Coyne, 2016). As long as promotion committees depend solely on traditional metrics for scientific productivity (Masum et al., 2013); as long as journal policies eschew giving credit for broadly collaborative, data-themed research (Nosek et al., 2015); as long as funding agencies relegate data sharing to the wish-list category of unfunded mandates (Royal Society Science Policy Centre, 2012); as long as easy-to-use platforms for uploading data remain isolated (Chismar, Horan, Hesse, Feldman, & Shaikh, 2011; Wicherts, Borsboom, Kats, & Molenaar, 2006); and as long as individual scientists fear being "scooped" when openly sharing hard-earned work to a field of "research parasites" (Longo & Drazen, 2016, p. 276; McKiernan et al., 2016), it will be difficult to break through the inertia of the status quo. Evidence has suggested that these constraints may be felt acutely by scientists early in their careers, for whom the stakes of a tenure-track publication record are highest (Tenopir et al., 2015). One compromise solution in this regard may be to encourage discov-

erability early by promoting the publication of metadata (i.e., data *about* the data, such as formats, variable names, and source) but then to allow for a reasonable embargo period so that the data generators can take full advantage of the value of their data sets before releasing them to the broader community. Another solution is to create platforms that easily support attribution of professional credit for contributing to data repositories. The NIMH's data archive was engineered to be such a platform, with digital object identifiers (DOIs) given to individual studies and data sets as a complement to the dois provided by publishers for published journal articles (Novikova, Hall, & Farber, 2017).

Fortunately, there is some evidence coming out of the natural experiments being conducted in other disciplines suggesting that the risk of change—at least under the right conditions—may be worth the movement forward. For example, in looking at over 10,000 studies concerning gene expression microarray data, Piwowar and Vision (2013) found a 9% overall citation advantage for those researchers who elected to share their data over those who did not, with a 30% citation advantage conferred for authors of older studies (published between 2004 and 2005) in their analysis pool after controlling for journal impact factor, number of authors, and other potential covariates. Other researchers found a 20% citation advantage for astronomy articles linking to open data sets, a 28%–50% citation advantage for astrophysics articles, and a 35% advantage for data-linked articles in paleoceanography (McKiernan et al., 2016). In the neurosciences, archivists reported success in gathering together more than 8,000 shared magnetic resonance imaging data sets by 2014. Communities are beginning to utilize these shared data as an opportunity to ask new scientific questions, to conduct integrative analyses, to test and improve the reproducibility of findings, to craft new and advanced analytic techniques (some relying on machine learning and other extrapolations of artificial intelligence engineering), and to conserve costs (Poldrack & Gorgolewski, 2014).

Nevertheless, even within the milieu of these early successes there is still reason to exercise caution. The dark side of open science has been a proliferation of spurious offers from publishers with specious track records (Betz, 2016). Ongoing, association-wide discussions about what constitutes scholarly excellence at a time when the frameworks for publication, as well as the metrics for professional contribution, are changing would help guide APA authors in evaluating the merit of open source publishing opportunities. Likewise, data breaches publicized in the national media—such as the 2013–2014 breach at Yahoo, the 2012–2014 breach of the U.S. Office of Personnel Management, and the 2016 ransomware attack on Hollywood Presbyterian Hospital—can have a chilling effect on how willing participants might be to share their personal information for long-term archiving in research (Elhai & Hall, 2016). Taking stock of the high-level security efforts being expended to protect personal information (Perakslis, 2014; Perakslis & Stanley, 2016), along with provisions for withdrawing consent if data integrity is in question (President's Council of Advisors on Science and Technology, 2010b), can go a long way in assuaging participants' concerns. Finally, for those areas in which data sharing might be turning into the norm, there will be a concomitant need to think through the infrastructure questions of how these data will be stored for community access in the long term, who will pay for that storage, and who will be able to benefit from these archives in the future (Pearce & Smith, 2011; Poldrack & Gorgolewski, 2014).

Academic training programs will need to adapt to this new world (Piwowar et al., 2008), especially as newer statistical methods are introduced to protect against an inflation of potential Type I errors as integrated samples increase the overall number of records available for post hoc review (Foster & Kalil, 2008). In fact, some have anticipated that one of the benefits of shared data might be an evolution of statistical thinking away from a constrained reliance on the artificial threshold of a $p < .05$ threshold to a more meta-analytic framework based on effect sizes or conditional probabilities for true positives–true negatives compared against probabilities for obtaining false positive–false negatives (Baldwin & Del Re, 2016; Dienes, 2016; Hripcsak et al., 2016).

Creating Better Pathways for Walking the Walk

How then can psychologists help create the pathways that will help move forward the ideal of an open, transparent environment for science within our own communities, while simultaneously applying the insights of psychological evidence to the task of catalyzing desirable change across all the sciences? I would like to think that there are least four simple things that the discipline can begin doing now to take the future in hand.

1. *Nurture innovation and agile learning.* One of the hallmarks of the information revolution is the ability to innovate in an agile way within a rapidly changing business environment and then from those iterations select best practice (Goldman, Nagel, & Preiss, 1995). To be sure, many of these rapid prototyping development practices occur within the proprietary boundaries of nondisclosure agreements and are not shared with others. Still, there is a growing swath of "agile" business models that are embracing many of the cultural tenets guiding the open science movement, not the least of which is an assumption that in the information age organizations perform better when they "cooperate to compete" (Preiss, Goldman, & Nagel, 1996, p. 1). Sparked by the success of information technology and biotechnology firms that routinely share technical information in community knowledge environments referred to as wikis (de-

rived from the Hawaiian word for "quick"), economists have begun to refer to these new collaborative business models as comprising a new type of business strategy some have referred to as "wikinomics" (Tapscott & Williams, 2008, p. 7).

Some behavioral researchers have argued that these newly emerging business principles could be applied to the scientific enterprise under the rubric of "agile science" (Hekler et al., 2016, p. 1). Indeed, many methodological experiments have been under way to improve the rapidity, the responsiveness, and the relevance of work conducted within the scientific enterprise using principles of rapid iteration and collaborative knowledge sharing (Riley, Glasgow, Etheredge, & Abernethy, 2013). Although other disciplines may have been ahead of the game in applying these techniques and in dealing with the computational opportunities associated with the resulting troves of "big data," the data technologies are indeed beginning to emerge that would allow scientists and statisticians to pursue the same approach in the social sciences (Hesse, Moser, & Riley, 2015). As the field learns from these natural experiments and as it shares its successes and failures—following an open science credo—it can adjust its own practices and policies accordingly.

2. *Think outside the paradigm.* When Thomas Kuhn described the concept of a paradigm shift in his seminal work on the structure of scientific revolutions, he explained how adopting a new conceptual framework for conducting normal science would necessarily lead to new ways of thinking (Kuhn, 1962). Engineers from Google illustrated this point aptly. Technologically, Google reproduces the links, data structures, and much of the content of the entire public-facing World Wide Web through automated Web crawlers and massive server farms on a routine basis. When the Google engineers offered to share these data with academic scientists, the proposals they received were narrow and reductionist in scope. The engineers had to readjust the academics' expectations, telling them to think anew about the questions they would ask if they had access to all the data in the world.

The point, even if a bit hyperbolic, is insightful. Most of the statistical protocols and embedded assumptions guiding the conduct of everyday psychological science were conceived during a period of paper-only publication, and because of space limitations, journal articles often lacked reports of effect size and uncertainty measures to go beyond the reporting of simple *p* values (Chavalarias, Wallach, Li, & Ioannidis, 2016). If we can begin to "walk the walk" of open science, the opportunities can begin to open up for conceiving of new statistical methods and new scientific questions more appropriately suited for a data-rich environment (Hey, Tansley, & Tolle, 2009; Riley, 2016). Bayesian techniques, for example, could be used to adjust the ways in which progress on theory is made in an open science environment by using new data to update a posteriori understanding of a theoretical relationship given the presence of new data (Dienes, 2016). In this sense, statistics such as the underused Bayes factor could help overcome the limitations of traditional null hypothesis testing by giving analysts a more nuanced ability to evaluate both the strength of H0 and H1 in comparison to each other in the light of data; that is, Bayes factor $= P(\mathrm{H1}|\mathrm{D})/P(\mathrm{H0}|\mathrm{D}) = P(\mathrm{D}/|\mathrm{H1}) \times P(\mathrm{D}|\mathrm{H0})$.

Even if implementation of a Bayes factor remains controversial, shared data environments could at least push for a more harmonized use of common metrics and ontologies to facilitate syntheses with future studies (Larsen et al., 2017). Meanwhile, collaborative platforms could be engineered to nudge authors toward including the metadata, the effects sizes, and uncertainty indicators needed to help with future meta-analyses (Tsuji, Bergmann, & Cristia, 2014). Machine-learning algorithms could make short work of traversing the millions of new connections available for comparison in big data research environments and thus accelerate the path from big data to knowledge (Hey et al., 2009). In one poignant example, a group of neuropsychological researchers reported running pattern-matching algorithms against a collectively stored archive of fMRI scans to uncover common features across subjects in just a fraction of the time it would have taken to look at each of these scans individually (Choudhury et al., 2014). Unapologetically, the goal of these paradigm-shifting discussions is to restructure the incentives and practices of science to promote a joint quest for scientific "truth over publishability" (Nosek, Spies, & Motyl, 2012, p. 615).

3. *Create simplicity from complexity.* When discussing obstacles to open collaboration, it is easy to recognize just how much work it often is to find an appropriate archive to secure the data for preservation, to deidentify the data to protect confidentiality, to protect intellectual property through legal reviews, and to transmit data assets for electronic archiving. Often, it is just more expedient to move on to the next pressing publication or proposal. It is for this reason that most of the successful data sharing efforts to date have made it a priority to simplify the process of moving personal contributions into a highly usable, easily accessible electronic platform referred to as a "data commons" (Stein, Adams, & Chambers, 2016, p. 1021). Nowhere is this more evident than in the commercial world, where participative contributions have become a staple in the sharing economy. Sites such as YouTube, SlideShare, LinkedIn, and FaceBook all make it as easy as possible to upload electronic resources into the public commons. In the vernacular of human factors specialists, they "reduce friction" in the user interface by making it easy to select and upload files for sharing.

In 2005, the open access advocate Linus Torvalds (originator of the open-source Linux Operating System) created

an open source tool called Git for maintaining version control in the electronic files by programmers in distributed software teams (Perkel, 2016). Statisticians and programmers from the sciences have begun taking advantage of the tool through a Web-based service called GitHub. GitHub has been useful in helping scientists to manage project flow, to maintain version control on common resources and data, and even to contribute statistical code for open evaluation and reuse by others (Perkel, 2016). In 2012, the Center for Open Science launched the reproducibility project to improve the replicability and rigor of findings in the psychological sciences. They used an open science framework based on principles embodied within GitHub to support team efforts to reproduce findings in psychology. The center's resulting collaborative platforms allowed users to register studies before data collection begins, share data for post hoc replication, split data (called forking) for parallel investigations, invite authors to join in the replication effort, and update data tables quickly as improvements were made to the underlying statistical code (Kubilius, 2014).

4. *Participate in continuous learning evidence platforms.* One area of particularly high stakes consideration for the psychological community is to stay involved in the continuous learning data platforms being put into place in health care to generate evidence for therapeutic practice. To support the many policy decisions needed for steering the health-care system at a time of excessive costs and rapid change, the connected data streams from electronic health records are being merged into a vision for a continuous learning evidence-based platform referred to by the Institute of Medicine as the "learning healthcare system" (Etheredge, 2007, p. 107; Grossmann, 2011, p. xii). Legislative and lobbying efforts have been under way to ensure that mental health and behavioral services get included in the mandate to utilize these data structures for the improvement of care (Cohen, 2015). These lobbying efforts should be important to the psychological community for at least three reasons: (a) tracking mechanisms for quantifying patients' mental health service needs can assist in resource allocation, (b) outcome assessments related to therapeutic effectiveness can serve as an evidentiary platform for comparative effectiveness research, and (c) the rich contextual and process data available through electronic health records can inform new scientific questions related to clinical treatment.

Although the concept of the learning health-care system may seem to be restricted to clinical practitioners and health psychologists, its principles may be generalizable to other areas of psychological discovery as well. Data systems infused within educational networks can provide a more systematic view of what works in educational psychology, whereas big data infused within operational business systems can lead to new insights on what works in the areas of industrial, organizational, and applied psychology. Even the once-restricted laboratories of experimental and social psychology have begun to take advantage of open network technologies by conducting virtual experimentation through crowdsourcing platforms. In May 2016, nearly 23,000 people voluntarily took part in thousands of social science experiments without ever visiting a lab, explained science writer Bohannon (2016, p. 1263). The studies used a crowdsourcing platform developed by the online retail giant Amazon called Mechanical Turk to solicit input from an online volunteer crowd for modest compensation. Some of these psychological experiments included a test on the limits of generosity, a comparison of religiosity and humility, and a measurement of the psychological impact from graphic warnings, to name just a few (Bohannon, 2016). Indeed, the demand for Internet platforms to crowdsource psychological science was so great that a group of scientists at New York University launched an open-source framework called psiTurk that could be used to conduct replicable behavioral experiments online in standardized ways (Gureckis et al., 2016).

Conclusion

In this article, I have tried to take an objective view of what data sharing might mean to the field of psychology using the tenets of a broadening open science culture. Candidly, the term *open science* can mean many things to many people. For some, it may represent a radically disruptive influence on the business of science wherein the "pay walls" of a fee-for-knowledge architecture come tumbling down and are replaced with new, innovative platforms for scientific participation akin to what may be happening in the world of transportation through Uber or the world of entertainment through YouTube. I am not discounting the importance of innovation in the methods of science; I absolutely believe that innovation is critical. I am, however, siding with the viewpoint expressed by the Royal Society Science Policy Centre in the United Kingdom that science should already be considered an inherently open enterprise (Royal Society Science Policy Centre, 2012). It is *how* we achieve that openness in an evolving culture influenced by scientific investments in globally distributed, shared data networks that is beginning to change the way our community can work together. We must begin to go through the hard work of changing our systems so that psychologists from around the globe can work collaboratively and substantively to solve some of the biggest global challenges confronting the world today, from mapping the human brain to treating mental illness and even to preventing and controlling diseases such as cancer. How we proceed will require a change in mind set and a lot more openness, including "open data, open collaboration, and above all, open minds" (Ong, 2016, p. 6).

References

21st Century Cures Act, H.R. 34, 114th Congress, (2015).

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., . . . Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27,* 108–119. http://dx.doi.org/10.1002/per.1919

Bainter, S. A., & Curran, P. J. (2015). Advantages of integrative data analysis for developmental research. *Journal of Cognition and Development, 16,* 1–10. http://dx.doi.org/10.1080/15248372.2013.871721

Baldwin, S. A., & Del Re, A. C. (2016). Open access meta-analysis for psychotherapy research. *Journal of Counseling Psychology, 63,* 249–260. http://dx.doi.org/10.1037/cou0000091

Bangdiwala, S. I., Bhargava, A., O'Connor, D. P., Robinson, T. N., Michie, S., Murray, D. M., . . . Pratt, C. A. (2016). Statistical methodologies to pool across multiple intervention studies. *Translational Behavioral Medicine, 6,* 228–235. http://dx.doi.org/10.1007/s13142-016-0386-8

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research, 41*(D1), D36–D42. http://dx.doi.org/10.1093/nar/gks1195

Betz, C. L. (2016). Authors beware: Open access predatory journals. *Journal of Pediatric Nursing, 31,* 233–234. http://dx.doi.org/10.1016/j.pedn.2016.02.006

Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., Claffey, M. P., . . . Patrick, K. (2016). Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association, 23*(e1), e42–e48. http://dx.doi.org/10.1093/jamia/ocv118

Bohannon, J. (2014, May 23). Replication effort provokes praise—And "bullying" charges. *Science, 344,* 788–789. http://dx.doi.org/10.1126/science.344.6186.788

Bohannon, J. (2016, June 10). Mechanical Turk upends social sciences. *Science, 352,* 1263–1264. http://dx.doi.org/10.1126/science.352.6291.1263

Buttliere, B. T. (2014). Using science and psychology to improve the dissemination and evaluation of scientific work. *Frontiers in Computational Neuroscience, 8,* 82. http://dx.doi.org/10.3389/fncom.2014.00082

California Institute for Telecommunications and Information Technology. (2014). *Personal data for the public good: New opportunities to enrich understanding of individual and population health*. Retrieved from http://hdexplore.calit2.net/wp-content/uploads/2015/08/hdx_final_report_small.pdf

Chandra, R., Iyer, R. S., & Raman, R. (2015). Enabling organizations to implement smarter, customized social computing platforms by leveraging knowledge flow patterns. *Journal of Knowledge Management, 19,* 95–107. http://dx.doi.org/10.1108/jkm-11-2014-0455

Chavalarias, D., Wallach, J. D., Li, A. H., & Ioannidis, J. P. (2016). Evolution of reporting *p* values in the biomedical literature, 1990–2015. *Journal of the American Medical Association, 315,* 1141–1148. http://dx.doi.org/10.1001/jama.2016.1952

Chavan, V., & Penev, L. (2011). The data paper: A mechanism to incentivize data publishing in biodiversity science. *Bioinformatics, 12*(Suppl. 15), S2. http://dx.doi.org/10.1186/1471-2105-12-S15-S2

Chismar, W., Horan, T. A., Hesse, B. W., Feldman, S. S., & Shaikh, A. R. (2011). Health cyberinfrastructure for collaborative use-inspired research and practice. *American Journal of Preventive Medicine, 40*(5, Suppl. 2), S108–S114. http://dx.doi.org/10.1016/j.amepre.2011.01.002

Choudhury, S., Fishman, J. R., McGowan, M. L., & Juengst, E. T. (2014). Big data, open science and the brain: Lessons learned from genomics. *Frontiers in Human Neuroscience, 8,* 239. http://dx.doi.org/10.3389/fnhum.2014.00239

Chung, K., Mullner, R., & Yang, D. (2002). Access to microdata on the Internet: Web-based analysis and data subset extraction tools. *Journal of Medical Systems, 26,* 555–560. http://dx.doi.org/10.1023/A:1020248827017

Cohen, D. (2015). Effect of the exclusion of behavioral health from health information technology (hit) legislation on the future of integrated health care. *Journal of Behavioral Health Services & Research, 42,* 534–539. http://dx.doi.org/10.1007/s11414-014-9407-x

Collins, F. S., & Riley, W. T. (2016). NIH's transformative opportunities for the behavioral and social sciences. *Science Translational Medicine, 8,* 366ed14. http://dx.doi.org/10.1126/scitranslmed.aai9374

Cooper, H. M., & VandenBos, G. R. (2013a). Archives of scientific psychology: A new journal for a new era. *Archives of Scientific Psychology, 1,* 1–6. http://dx.doi.org/10.1037/arc0000001

Cooper, H. M., & VandenBos, G. R. (2013b). *Ten prescriptions to foster better social science in the internet environment: Final report on NIH grant*. Unpublished Report, Duke University.

Coyne, J. C. (2016). Replication initiatives will not salvage the trustworthiness of psychology. *BMC Psychology, 4,* 28. http://dx.doi.org/10.1186/s40359-016-0134-3

Crane, D. (1972). *Invisible colleges; diffusion of knowledge in scientific communities*. Chicago, IL: University of Chicago Press.

Critchley, C. R., Bruce, G., & Farrugia, M. (2013). The impact of commercialisation on public perceptions of stem cell research: Exploring differences across the use of induced pluripotent cells, human and animal embryos. *Stem Cell Reviews, 9,* 541–554. http://dx.doi.org/10.1007/s12015-013-9445-4

Cuthbert, B. N., & Kozak, M. J. (2013). Constructing constructs for psychopathology: The NIMH research domain criteria. *Journal of Abnormal Psychology, 122,* 928–937. http://dx.doi.org/10.1037/a0034028

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An assessment of published and nonpublished correlation matrices. *Personnel Psychology, 65,* 221–249. http://dx.doi.org/10.1111/j.1744-6570.2012.01243.x

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72,* 78–89. http://dx.doi.org/10.1016/j.jmp.2015.10.003

Elhai, J. D., & Hall, B. J. (2016). Anxiety about internet hacking: Results from a community sample. *Computers in Human Behavior, 54,* 180–185. http://dx.doi.org/10.1016/j.chb.2015.07.057

Engelbart, D. C. (1961). *Augmenting human intellect: A conceptual framework*. Retrieved from https://www.princeton.edu/~hos/h598/augmenting.pdf

Etheredge, L. M. (2007). A rapid-learning health system. *Health Affairs, 26*(2), w107–w118. http://dx.doi.org/10.1377/hlthaff.26.2.w107

Fair Access to Science and Technology Research Act (FASTR), S.350, 113th Congress (2013).

Fair Access to Science and Technology Research Act (FASTR), S.779, 114th Congress (2015).

Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In S. Bartling & S. Friesike (Eds.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing* (pp. 17–47). http://dx.doi.org/10.1007/978-3-319-00026-8_2

Foster, E. M., & Kalil, A. (2008). New methods for new questions: Obstacles and opportunities. *Developmental Psychology, 44,* 301–304. http://dx.doi.org/10.1037/0012-1649.44.2.301

Giles, J. (2005, December 15). Internet encyclopaedias go head to head. *Nature, 438,* 900–901. http://dx.doi.org/10.1038/438900a

Goldman, S. L., Nagel, R. N., & Preiss, K. (1995). *Agile competitors and virtual organizations: Strategies for enriching the customer*. New York, NY: Van Nostrand Reinhold.

Gorgolewski, K. J., Margulies, D. S., & Milham, M. P. (2013). Making data sharing count: A publication-based approach. *Frontiers in Neuroscience, 7,* 9. http://dx.doi.org/10.3389/fnins.2013.00009

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., . . . Margulies, D. S. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics, 9,* 8. http://dx.doi.org/10.3389/fninf.2015.00008

Grossmann, C. (2011). *Engineering a learning healthcare system: A look at the future: Workshop summary*. Washington, DC: National Academies Press.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., . . . Chan, P. (2016). psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods, 48,* 829–842. http://dx.doi.org/10.3758/s13428-015-0642-8

Handler, R. A., & Ferrer Conill, R. (2016). Open data, crowdsourcing and game mechanics: A case study on civic participation in the digital age. *Computer Supported Cooperative Work, 25,* 153–166. http://dx.doi.org/10.1007/s10606-016-9250-0

Hedrick, T. E. (1988). Justifications for the sharing of social science data. *Law and Human Behavior, 12,* 163–171. http://dx.doi.org/10.1007/BF01073124

Hekler, E. B., Klasnja, P., Riley, W. T., Buman, M. P., Huberty, J., Rivera, D. E., & Martin, C. A. (2016). Agile science: Creating useful products for behavior change in the real world. *Translational Behavioral Medicine, 6,* 317–328. http://dx.doi.org/10.1007/s13142-016-0395-7

Hesse, B. W., Hansen, D., Finholt, T., Munson, S., Kellogg, W., & Thomas, J. C. (2010). Social participation in Health 2.0. *Computer, 43,* 45–52. http://dx.doi.org/10.1109/MC.2010.326

Hesse, B. W., Moser, R. P., & Riley, W. T. (2015). From big data to knowledge in the social sciences. *Annals of the American Academy of Political and Social Science, 659,* 16–32. http://dx.doi.org/10.1177/0002716215570007

Hesse, B. W., Sproull, L., Kiesler, S. B., & Walsh, J. P. (1993). Returns to science: Computer networks in oceanography. *Communications of the ACM, 36,* 90–101. http://dx.doi.org/10.1145/163381.163409

Hey, A. J. G., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research.

Holdren, J. P. (2013). *Memorandum for the heads of executive departments and agencies, 22 February, 2013*. Washington, DC: Executive Office of the President. Retrieved from https://obamawhitehouse.archives.gov

Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V., . . . Madigan, D. (2016). Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences of the United States of America, 113,* 7329–7336. http://dx.doi.org/10.1073/pnas.1510502113

Hudson, K. L., & Collins, F. S. (2017). The 21st Century Cures Act—A view from the NIH. *New England Journal of Medicine, 376,* 111–113. http://dx.doi.org/10.1056/NEJMp1615745

Institute of Medicine. (2008). *Cancer care for the whole patient: Meeting psychosocial health needs*. http://dx.doi.org/10.17226/11993

Institute of Medicine. (2013). *Best care at lower cost: The path to continuously learning health care in America*. http://dx.doi.org/10.17226/13444

Institute of Medicine. (2015). *Capturing social and behavioral domains and measures in electronic health records: Phase 2*. http://dx.doi.org/10.17226/18951

Ioannidis, J. P. (2016). Why most clinical research is not useful. *PLoS Medicine, 13*(6), e1002049. http://dx.doi.org/10.1371/journal.pmed.1002049

Joly, Y., Dalpé, G., So, D., & Birko, S. (2015). Fair shares and sharing fairly: A survey of public views on open science, informed consent and participatory research in biobanking. *PLoS ONE, 10*(7), e0129893. http://dx.doi.org/10.1371/journal.pone.0129893

Kubilius, J. (2014). Sharing code. *i-Perception, 5,* 75–78. http://dx.doi.org/10.1068/i004ir

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Landis, D., Courtney, W., Dieringer, C., Kelly, R., King, M., Miller, B., . . . Calhoun, V. D. (2016). COINS data exchange: An open platform for compiling, curating, and disseminating neuroimaging data. *NeuroImage, 124*(Pt. B), 1084–1088. http://dx.doi.org/10.1016/j.neuroimage.2015.05.049

Larsen, K. R., Michie, S., Hekler, E. B., Gibson, B., Spruijt-Metz, D., Ahern, D. K., . . . Yi, J. (2017). Behavior change interventions: The potential of ontologies for advancing science and practice. *Journal of Behavioral Medicine, 40,* 6–22. http://dx.doi.org/10.1007/s10865-016-9768-0%P

Longo, D. L., & Drazen, J. M. (2016). Data sharing. *New England Journal of Medicine, 374,* 276–277. http://dx.doi.org/10.1056/NEJMe1516564

Masum, H., Rao, A., Good, B. M., Todd, M. H., Edwards, A. M., Chan, L., . . . Bourne, P. E. (2013). Ten simple rules for cultivating open science and collaborative R&D. *PLoS Computational Biology, 9*(9), e1003244. http://dx.doi.org/10.1371/journal.pcbi.1003244

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., . . . Yarkoni, T. (2016). Point of view: How open science helps researchers succeed. *eLife, 5,* e16800. http://dx.doi.org/10.7554/eLife.16800

McLuhan, M., Gordon, W. T., Lamberti, E., & Scheffel-Dunand, D. (2011). *The Gutenberg galaxy: The making of typographic man* (1st ed.). Toronto, Ontario, Canada: University of Toronto Press.

Michie, S., Carey, R. N., Johnston, M., Rothman, A. J., de Bruin, M., Kelly, M. P., & Connell, L. E. (2016). From theory-inspired to theory-based interventions: A protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. *Annals of Behavioral Medicine*. Advance online publication. http://dx.doi.org/10.1007/s12160-016-9816-6

National Academy of Sciences. (1989). *Information technology and the conduct of research: The user's view*. http://dx.doi.org/10.17226/763

National Committee on Vital and Health Statistics. (2001). *Information for health: A strategy for building the national health information infrastructure*. Retrieved from https://aspe.hhs.gov/report/information-health-strategy-building-national-health-information-infrastructure

National Institutes of Health. (2016). *NIH public access policy details*. Retrieved from https://publicaccess.nih.gov/policy.htm

Nelson, B. (2009). Data sharing: Empty archives. *Nature, 461,* 160–163. http://dx.doi.org/10.1038/461160a

Nielsen, M. A. (2012). *Reinventing discovery: The new era of networked science*. Princeton, NJ: Princeton University Press.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015, June 26). Promoting an open research culture. *Science, 348,* 1422–1425. http://dx.doi.org/10.1126/science.aab2374

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7,* 615–631. http://dx.doi.org/10.1177/1745691612459058

Novikova, S., Hall, D., & Farber, G. K. (2017). The s-index and the National Institute of Mental Health data archive. *Journal of the American Medical Association Psychiatry, 74,* 540–541. http://dx.doi.org/10.1001/jamapsychiatry.2017.0145

Ong, M. B. H. (2016). Biden designates nci's genomic data commons as foundation of moonshot. *The Cancer Letter, 42,* 1–8.

Park, T. (2011). Information "liberación." *Healthcare Informatics, 28,* 45–46.

Pearce, N., & Smith, A. H. (2011). Data sharing: Not as simple as it seems. *Environmental Health, 10,* 107. http://dx.doi.org/10.1186/1476-069X-10-107

Perakslis, E. D. (2014). Cybersecurity in health care. *New England Journal of Medicine, 371,* 395–397. http://dx.doi.org/10.1056/NEJMp1404358

Perakslis, E. D., & Stanley, M. (2016). A cybersecurity primer for translational research. *Science Translational Medicine, 8,* 322ps2. http://dx.doi.org/10.1126/scitranslmed.aaa4493

Perkel, J. (2016, October 28). Democratic databases: Science on GitHub. *Nature, 538,* 127–128. http://dx.doi.org/10.1038/538127a

Pirolli, P., Preece, J., & Shneiderman, B. (2010). Cyberinfrastructure for social action on national priorities. *Computer, 43,* 20–21. http://dx.doi.org/10.1109/mc.2010

Piwowar, H. A., Becich, M. J., Bilofsky, H., & Crowley, R. S. (2008). Towards a data sharing culture: Recommendations for leadership from academic health centers. *PLoS Medicine, 5*(9), e183. http://dx.doi.org/10.1371/journal.pmed.0050183

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ, 1:* e175. http://dx.doi.org/10.7717/peerj.175

Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: Data sharing in neuroimaging. *Nature Neuroscience, 17,* 1510–1517. http://dx.doi.org/10.1038/nn.3818

Preece, J. (2016). Citizen science: New research challenges for human-computer interaction. *International Journal of Human-Computer Interaction, 32,* 585–612. http://dx.doi.org/10.1080/10447318.2016.1194153

Preiss, K., Goldman, S. L., & Nagel, R. N. (1996). *Cooperate to compete: Building agile business relationships*. New York, NY: Van Nostrand Reinhold.

President's Council of Advisors on Science and Technology. (2010a). *Designing a digital future: Federally funded research and development in networking and information technology*. Washington, DC: Executive Office of the President.

President's Council of Advisors on Science and Technology. (2010b). *Realizing the full potential of health information technology to improve healthcare for Americans: The path forward*. Washington, DC: Executive Office of the President.

Reid, A. T., Bzdok, D., Genon, S., Langner, R., Muller, V. I., Eickhoff, C. R., . . . Eickhoff, S. B. (2016). Anima: A data-sharing initiative for neuroimaging meta-analyses. *Neuroimage, 124*(Pt. B), 1245–1253. http://dx.doi.org/10.1016/j.neuroimage.2015.07.060

Riley, W. T. (2016). *A new era of clinical research in a data-rich environment*. In B. W. Hesse, D. K. Ahern, & E. Beckjord (Eds.), Oncology informatics: Using health information technology to improve processes and outcomes in cancer (pp. 343–355). http://dx.doi.org/10.1016/B978-0-12-802115-6.00018-5

Riley, W. T., Glasgow, R. E., Etheredge, L., & Abernethy, A. P. (2013). Rapid, responsive, relevant (R3) research: A call for a rapid learning health research enterprise. *Clinical and Translational Medicine, 2,* 10. http://dx.doi.org/10.1186/2001-1326-2-10

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641. http://dx.doi.org/10.1037/0033-2909.86.3.638

Royal Society Science Policy Centre. (2012). *Science as an open enterprise*. London, United Kingdom: Author.

Shneiderman, B. (2011, July). *Technology-mediated social participation: The next 25 years of human computer interaction challenges*. Paper presented at the 14th International Conference on Human-Computer Interaction, Orlando, FL.

Sproull, L., & Kiesler, S. (1991). *Connections: New ways of working in the networked organization*. Cambridge, MA: MIT Press.

Stein, B. D., Adams, A. S., & Chambers, D. A. (2016). A learning behavioral health care system: Opportunities to enhance research. *Psychiatric Services, 67,* 1019–1022. https://doi.org/10.1176/appi.ps.201500180

Tapscott, D., & Williams, A. D. (2008). *Wikinomics: How mass collaboration changes everything* (Expanded ed.). New York, NY: Portfolio.

Tardif, C. L., Schafer, A., Trampel, R., Villringer, A., Turner, R., & Bazin, P. L. (2016). Open science CBS neuroimaging repository: Sharing ultra-high-field MR images of the brain. *Neuroimage, 124*(Pt. B), 1143–1148. http://dx.doi.org/10.1016/j.neuroimage.2015.08.042

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE, 6*(6), e21101. http://dx.doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., . . . Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE, 10*(8), e0134826. http://dx.doi.org/10.1371/journal.pone.0134826

Topol, E. J. (2013). *The creative destruction of medicine: How the digital revolution will create better health care*. New York, NY: Basic Books.

Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analyses: Toward cumulative data assessment. *Perspectives on Psychological Science, 9,* 661–665. http://dx.doi.org/10.1177/1745691614552498

Van Horn, J. D., & Gazzaniga, M. S. (2002). Databasing fMRI studies—Towards a "discovery science" of brain function. *Nature Reviews Neuroscience, 3,* 314–318. http://dx.doi.org/10.1038/nrn788

Wang, L., Alpert, K. I., Calhoun, V. D., Cobia, D. J., Keator, D. B., King, M. D., . . . Ambite, J. L. (2016). Schizconnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration. *NeuroImage, 124*(Pt. B), 1155–1167. http://dx.doi.org/10.1016/j.neuroimage.2015.06.065

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61,* 726–728. http://dx.doi.org/10.1037/0003-066X.61.7.726

Wood, D., King, M., Landis, D., Courtney, W., Wang, R., Kelly, R., . . . Calhoun, V. D. (2014). Harnessing modern web application technology to create intuitive and efficient data visualization and sharing tools. *Frontiers in Neuroinformatics, 8,* 71. http://dx.doi.org/10.3389/fninf.2014.00071

Woolley, J. P., McGowan, M. L., Teare, H. J. A., Coathup, V., Fishman, J. R., Settersten, R. A., Jr., . . . Juengst, E. T. (2016). Citizen science or scientific citizenship? Disentangling the uses of public engagement rhetoric in national research initiatives. *BMC Medical Ethics, 17,* 33. http://dx.doi.org/10.1186/s12910-016-0117-1

Received February 8, 2017
Revision received July 20, 2017
Accepted July 21, 2017 ■