Data Sharing and the Growth of Medical Knowledge
Annette Flanagin, RN, MA; Gregory Curfman, MD; Kirsten Bibbins-Domingo, PhD, MD, MAS

In medical research, data sharing facilitates discovery and innovation, transparency, and reproducibility, and, ultimately, trust in science. Impelled by the COVID-19 pandemic, demands for data sharing have accelerated with increasing calls for more rapid dissemination, assessment, combination, and analyses of new medical research results.

Contemporary recommendations for data sharing are based on policies developed 4 decades ago. For example, GenBank was established in 1982 as a public access repository of nucleotide sequences.1 Now maintained by the US National Center for Biotechnology Information, GenBank contains more than 240 million publicly available sequences.2 In 1985, the US National Research Council (NRC)3 released a report on data sharing that continues to serve as a useful guide for researchers, authors, editors, and journals. Among the NRC’s recommendations, the following have relevance for scientific journal publication.3,4

Data sharing should be a regular practice. Investigators should share data by the time of publication, except in compelling circumstances.

Plans for data sharing should be an integral part of a research plan whenever data sharing is feasible.

Subsequent researchers who request data from others should bear the associated incremental costs, should endeavor to keep the burdens of data sharing to a minimum, and should explicitly acknowledge the contribution of the initial investigators.

Journal editors should require authors to provide access to data as a part of the peer review process, give more emphasis to reports of secondary analyses and replications, require full credit and appropriate citations to original data collections in reports based on secondary analyses, and encourage authors to make detailed data accessible to other researchers.

Despite these early recommendations, policies encouraging data sharing have varied widely among research institutions, governmental agencies, funders, and journals, and it has taken 4 decades for policies of encouragement to become requirements. Journals have taken on various policies to encourage data sharing across different data and study types. Discussion and debate have revolved around whether enforcement of data sharing is within the responsibilities for journals, particularly when requirements for data sharing may be viewed as unfunded mandates and that the manner for sharing data is not universally accepted.

In 1989 PNAS was among the first journals to encourage depositing of genetic data sets in public repositories,5 and many journals (including JAMA) have followed this lead and encouraged authors to deposit genomic data sets (eg, protein or DNA sequences) in a publicly available database. Among medical journals, Annals of Internal Medicine6 and the BMJ7 began requiring authors of research articles to provide data availability or sharing statements in 2007 and 2009, respectively, and in 2014 the PLOS journals began to require “authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.”8 The Institute of Medicine, now National Academy of Medicine (in 2015)9 and the International Committee of Medical Journal Editors (in 2016 and 2017)10 announced similar calls for data sharing statements to be published with reports of clinical trials, and the journals following the ICMJE policy (including JAMA and the JAMA Network journals) made this a condition of publication for clinical trials. Although these policies encourage data sharing and require the publication of a data sharing statement, none require or enforce the actual sharing of data.

In 2003, the US National Institutes of Health (NIH) issued a statement that encouraged data sharing in general and required investigators submitting applications for grants of $500 000 or more in any single year to include a plan for data sharing or indicate why data sharing was not possible.11 Updates and stronger NIH recommendations ensued in subsequent years, including a policy requiring data sharing for NIH-funded clinical trials in 2016.12 In October 2020, the NIH announced that effective January 25, 2023, NIH-funded researchers must “prospectively plan for how scientific data will be preserved and shared through submission of a Data Management and Sharing Plan” and established “the expectation for maximizing the appropriate sharing of scientific data generated from NIH-funded or conducted research, with justified limitations or exceptions.”13

To facilitate the reporting of such data management and sharing plans, JAMA and the JAMA Network journals have expanded the requirement for a data sharing statement to authors of all research articles, not just reports of clinical trials.14 As of October 2022, at the time of manuscript submission of any research article, authors are asked a series of questions regarding their plans to share data. Authors are asked to identify the data that will be made available to others, including individual patient data, a data dictionary that defines each field in the data set, and supporting documentation (eg, statistical/analytic code); when, where, and how the data will be available (eg, a link to a data repository); types of analyses that are permitted; and if there will be any restrictions on the use of the data. Authors also have the option to explain why data may not be shared. While the JAMA Network journals endorse the principle of data sharing and have adopted this policy to encourage the sharing of data,
sharing is not mandated by the JAMA Network. An author’s intention to share data or not will not be considered in the editorial decision. The JAMA Network journals will continue to publish data sharing statements in an online supplement linked to the published article.

We believe that data sharing promotes the growth of medical knowledge. We also understand that some researchers may wish to guard their data or may not have the resources to effectively share their data. However, as mandates from funders and others continue to proliferate and strengthen, we hope the JAMA Network journal data sharing statement policy and process will enable and encourage the wider sharing of data by researchers and thereby promote discovery, transparency, reproducibility, and trust in medical science.

ARTICLE INFORMATION
Author Affiliations: Ms Flanagin is Executive Managing Editor, Dr Curfman is Interim Executive Editor, and Dr Bibbins-Domingo is Editor in Chief, JAMA and JAMA Network.
Corresponding Author: Annette Flanagin, RN, MA, JAMA and the JAMA Network (annette.flanagin@jamanetwork.org).
Published Online: December 5, 2022. doi:10.1001/jama.2022.22837
Conflict of Interest Disclosures: None reported.
REFERENCES