



The NEW ENGLAND JOURNAL of MEDICINE

Perspective

Avoiding Data Dumpsters — Toward Equitable and Useful Data Sharing

Laura Merson, Oumar Gaye, M.D., Ph.D., and Philippe J. Guerin, M.D., Ph.D.

The potential health benefits from sharing participant-level clinical research data for the purpose of secondary analysis or meta-analysis have been widely touted. Although some researchers

remain wary about sharing data, recent policies and proposals by funders, scientific journals, research institutions, and international health organizations mean that data sharing, in one form or another, is inevitable. Now is therefore the time to focus on developing practices for data sharing that are effective, efficient, equitable, and ethical. In the process, we may need to question the assumption that more is better. Simply making more data openly available may not lead to analyses that are relevant and that are actually applied to improve health.

A variety of data-sharing platform models have evolved to meet the needs of various communities. As more partners in science man-

date sharing of data, these platforms and repositories are likely to grow rapidly in number and size. But they will also need to evolve to avoid perils that could undermine the benefits of data sharing.

One of the risks posed by these expanding repositories is the production of “data dumpsters”: repositories of data without the metadata, data dictionaries, or documentation needed for meaningful or correct reanalysis. Fulfilling an obligation to share data before good practices in data formatting and documentation have been established and replicated may allow researchers to check the “data shared” box, but it may also result in an epidemic of ac-

cessible data of limited usefulness. There is currently inadequate funding and expertise for curating data to a standard and quality suitable for external secondary use; researchers must bear the costs themselves or opt, as many currently do, to make raw data available without the explanatory documentation necessary to make them useful. Most repositories are not equipped to rectify this problem — nor do they see this function as part of their mandate.

Another concern is the risk of widening the research-output gap between low-resource and high-resource countries. Analysts in rich countries have the skills and resources to use and reanalyze data collected in lower-income countries, whereas the reverse is rarely true. When medical journals mandate data sharing, researchers in low-income countries will have no choice but to allow external access to those who are better

equipped to make use of the data. But better equipped does not mean better qualified: if there's no requirement to involve primary researchers when conducting secondary analyses, misinterpretation of the data is possible — indeed it is likely, especially in the case of data sets for which high-quality data management and descriptors are lacking. Re-use of data that produces incorrect results does not improve health outcomes.

More investment is needed in platforms that can standardize, clean, and curate data into the usable formats that are required for sharing data effectively. Those systems will also have to ensure ethical and responsible data sharing that maximizes the use of available data. In global health, that means encouraging engagement from researchers around the world and ensuring appropriate acknowledgment of the data generators.

One proof of concept is provided by the WorldWide Antimalarial Resistance Network (www.WWARN.org), established by malaria researchers in 2009 to provide the evidence required to elucidate factors affecting the efficacy of antimalarial drugs, with a particular focus on resistance. The WWARN approach has been to facilitate collaborative study groups to answer specific research questions using pooled analyses of individual-participant data from multiple clinical trials. WWARN has had considerable

success in engaging a heterogeneous malaria research community; the network now comprises more than 260 collaborators in 70 countries, and our data repository contains the large majority of clinical trial data on current antimalarials generated by academic groups and the pharmaceutical industry. The benefits of pooling data are clear, and the outcomes have been palpable: by increasing sample sizes, researchers have identified trends or subpopulation effects with greater certainty, and their findings have led to changes in global treatment guidelines.

We believe the success of this network lies in its symbiotic approach — an approach that evolved out of necessity to address the fears of researchers in malaria-endemic countries that data would be “scooped” by analysts with greater resources for mining their potential. WWARN data generators are encouraged to be fully involved in the process of any meta-analysis, and data generators' contributions to any resulting publications are recognized in accordance with the guidelines on authorship provided by the International Committee of Medical Journal Editors and MEDLINE. WWARN adds value to the data by harmonizing heterogeneous data and metadata, thereby ensuring that all data generators can more easily analyze one another's data and encouraging maximum use to improve treatment outcomes.

The model established for malaria is working well and may serve as a prototype for other neglected tropical diseases, for which available data are even more limited. Such data are costly and difficult to collect, and because they have limited commercial interest, studies are often taxpayer-funded. Moreover, most of the knowledge generated by these studies will be of greatest use to lower-income countries. All these features highlight the ethical as well as economic imperative to share data from clinical trials of neglected diseases in ways that are efficient and equitable. A newly established Infectious Diseases Data Observatory (www.IDDO.org) expects to expand the principles and practices established by the malaria community to other infectious diseases, including visceral leishmaniasis and schistosomiasis. Our primary aim is to support scientific communities in data sharing that is truly useful and that produces new knowledge that is used to change lives.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the WorldWide Antimalarial Resistance Network, Infectious Diseases Data Observatory, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK (L.M., P.J.G.); and the Laboratory of Parasitology–Mycology, Faculty of Medicine, University Cheikh Anta Diop, Dakar, Senegal (O.G.).

This article was published on May 11, 2016, at NEJM.org.

DOI: 10.1056/NEJMp1605148

Copyright © 2016 Massachusetts Medical Society.