

Description

In presiding over the “cancer moon-shot,” Former Vice-President of the United States Joe Biden described data sharing as being essential to the advancement of biomedical research. Sharing research data is something the NIH supports, too, as the public interest in open data grows. During the research cycle, extensive datasets are generated and scientists stand to benefit—and *possibly suffer*—from data sharing plans established by funders and publishers. Data, the raw facts acquired during studies and trials, are the foundation of research and translational medicine**.** At its most basic level, the sharing of data refers to the accessibility of data by the public, since it was the public that originally funded such research. For this reason, it makes sense that the NIH data sharing policy reflects the larger shift in policies for researchers regarding data sharing practices. Indeed, clinicians and patients require data to be available if they wish to make the most informed decisions.

Data Sharing: The NIH and Beyond

Given the rapid shift in attitudes among funders and institutions in sharing research data, the NIH has created policies that require data sharing. This NIH Data Sharing Policy includes the sharing of data generated by studies with direct costs of more than $500,000 per year. Of course, there are exceptions, including patient protection guaranteed under governmental privacy rules. However, barring reasonable reasons to not share, data generated during the research cycle must be accessible following acceptance for publication of manuscripts associated with final data sets.

The NIH is not alone in requiring or strongly advocating for data sharing in research. For example, the Nature Publishing Group requires certain data sets, such as transcription profile arrays, to be deposited in a public database for later access. Similarly, the Medical Research Council in the United Kingdom (UK) has policies in place to maximize research productivity by making data available for use by other researchers. Meanwhile, the International Committee of Medical Journal Editors (ICMJE) has created a policy in which de-identified individual data points from clinical trials must be made available within six months of publication. Such policies, while laudable for their desire to accelerate discovery, do not come without criticism. The ICMJE policy of six months is argued to be too short since many investigators conduct secondary studies or perform additional analyses with original clinical trial datasets. Along with these efforts to institute the sharing of research data, concerns over data sharing as a compulsory policy for researchers has grown.

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

Lessons from Sharing Data

There are many reported benefits of data sharing. For example, researchers having access to pre-existing data may then re-examine such data sets and possibly generate new insights. Nevertheless, there are several risks associated with sharing data – and researchers must be wary. Data access by individuals not originally involved in the associated study can result in the inappropriate use of such data. For example, researchers not involved in the original research may misinterpret findings due to a lack of understanding of the context of data generation. More sinister, however, is the prospect of the emergence of "research parasites." Such researchers may use another group's data for their own purposes, including scooping the future research plans of the original group or attempting to sabotage research competitors.

Conversely, symbiotic relationships must be established to ensure the productive use of deposited data. The Research Councils UK have provided guidelines that include being mindful of the intellectual contributions of those researchers who are involved in the generation of original data. As an early stage researcher, one must be wary of data that are made publicly available. Researchers must consider the potential consequences of depositing data into public repositories and consider the potential long-term uses of such data.

Open Data and Confidentiality

In an era of increasing publicly available datasets, what are other scientists doing? A survey conducted by Wiley of 90,000 authors of manuscripts related to health, life physical, and social sciences as well as humanities assessed the data sharing practices, attitudes, and motivations of researchers. In doing so, they found that 52% of researchers shared data. Among those sharing research data, two-thirds did so in the form of supplementary material in a journal. However well intending a journal is, research data deposited as supplementary data is not the best way to curate data generated during the research cycle. It is not easy for a scientist to download data sets behind pay walls or included as a PDF. Researchers should consider instead using public repositories, such as those suggested by the Nature Publishing Group, to maximize the efficacy of their public data.

It should be noted, however, that these data sharing practices varied by research discipline: 66% of life scientists shared data whereas only 36% of social scientists participated in data sharing. What was the main reason to not share data? Researchers described being hesitant due to intellectual property or confidentiality issues associated with their work. Furthermore, when data sharing policies for researchers were not in place, scientists simply would not share data.

As scientists continue to accelerate discovery and generate new possibility for humankind, the research cycle has evolved as well. Data sharing is increasingly becoming the norm, with major funding bodies and journals creating policies for researchers to promote data sharing. The NIH data sharing plan aims to facilitate the dissemination of raw data, but researchers must be careful with their data sharing practices. As science moves toward a future that is more open, careful collaborations are necessary to protect both research integrity and the researchers' interests as well.

Category

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

1. Publishing Research
2. Understanding Ethics

Date Created
2017/04/25
Author
daveishan

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license