Challenges and Opportunities

SCIENTIFIC INNOVATION HAS BEEN CALLED ON TO SPUR ECONOMIC recovery; science and technology are essential to improving public health and welfare and to inform sustainability; and the scientific community has been criticized for not being sufficiently accountable and transparent. Data collection, curation, and access are central to all of these issues. For this reason, *Science* has joined with colleagues from our sister publications *Science Signaling*, *Science Translational Medicine*, and *Science* Careers to provide a broad look at the issues surrounding the increasingly huge influx of research data. The entire collection is compiled online at www.sciencemag.org/special/data/. As you will discover, two themes appear repeatedly: Most scientific disciplines are finding the data deluge to be extremely challenging, and tremendous opportunities can be realized if we can better organize and access the data.

Our authors explore data issues that apply to specific fields as well as challenges shared between fields. These articles clearly show that the challenges are difficult and growing. We have recently passed the point where more data is being collected than we can physically store (see Hilbert *et al.,* published online). This storage gap will widen rapidly in data-intensive fields. Thus, decisions will be needed on which data to archive and which to discard. A separate problem is how to access and use these data. Many data sets are becoming too large to download. Even fields with well-established data archives, such as genomics, are facing new and growing challenges in data volume and management. And even where accessible, much data in many fields is too poorly organized to enable it to be efficiently used.

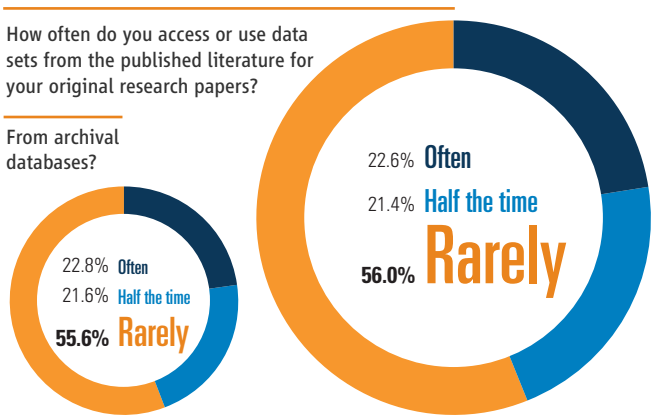
To delve deeper into these issues, *Science* polled our peer reviewers from last year about the availability and use of data. We received about 1700 responses, representing input from an international and interdisciplinary group of scientific leaders. About 20% of the respondents regularly use or analyze data sets exceeding 100 gigabytes, and 7% use data sets exceeding 1 terabyte. About half of those polled store their data only in their laboratories—not an ideal long-term solution. Many bemoaned the lack of common metadata and archives as a main impediment to using and storing data, and most of the respondents have no funding to support archiving.

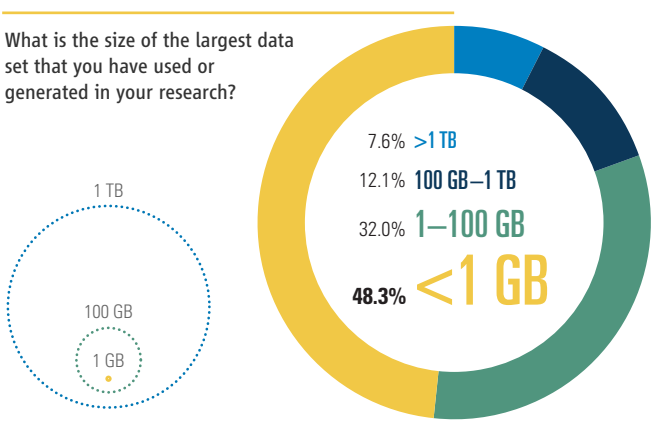
Many of the responders indicated that they seek or would like additional help in analyzing the data that they had collected. If we can use and reuse scientific data better, the opportunities, as indicated in many examples in this special section, are myriad. Large integrated data sets can potentially provide a much deeper understanding of both nature and society and open up many new avenues of research. And they are critical for addressing key societal problems—from improving public health and managing natural resources intelligently to designing better cities and coping with climate change.

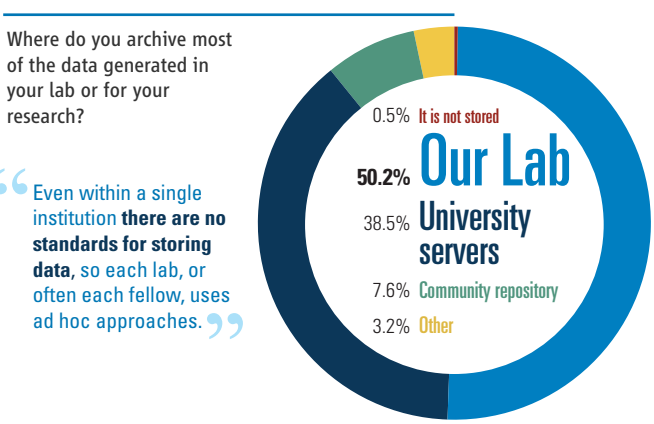
To realize these opportunities, many of the articles in this collection speak of changing the culture of science and the practices of scientists, as well as recognizing the growing responsibility for much better data stewardship. Several of the pieces illustrate steps toward these goals. But it is clear that organized effort and leadership are needed from funders, societies, journals, educators, and individual scientists—and from society at large.

We hope that this collection spurs additional thinking and catalyzes new efforts in dealing with these critical issues. As a start, we invite you to share your thoughts at talk.sciencemag.org, where you can also contribute to our poll.

– *SCIENCE* STAFF





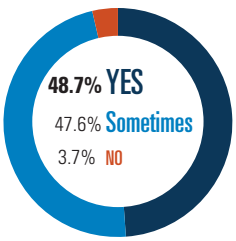


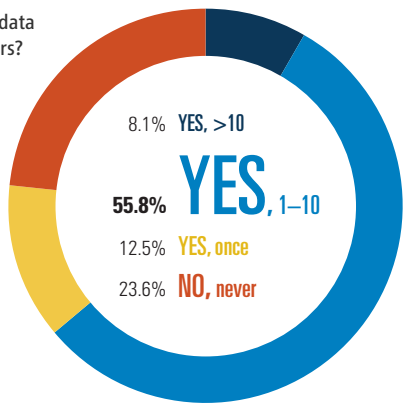
CREDIT: M. TWOMBLY/*SCIENCE*; SOURCE: *SCIENCE* ONLINE SURVEY

Published by AAAS

Have you asked colleagues for data related to their published papers?

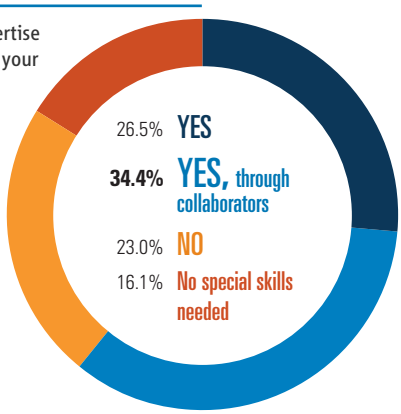
If you answered yes, have the appropriate data been provided?





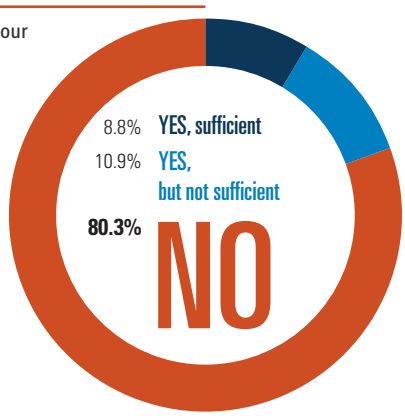
Do you have the necessary expertise in your lab or group to analyze your data in the way you want?

"The next few years [particularly in medicine] **the volume of data we need to analyze will expand exponentially.**"



Is there sufficient funding for your lab or research group for data curation?

"There are many tales of early archaeologists burning wood from the ruins to make coffee. If we fail to curate the environmental archives **we collect from nature at public expense**, we essentially repeat those mistakes."



CREDIT: M. TWOMBLY/*SCIENCE*; SOURCE: *SCIENCE* ONLINE SURVEY

CONTENTS

News

Perspectives

See also:

Editorial

News Focus

Books

Policy Forum

Science *Express Research Article**

Science Signaling*

Science Translational Medicine*

Science *Careers**

**These items, plus a related podcast and online discussion, are available at www.sciencemag.org/special/data/*

Published by AAAS