Consensus Study Report
HIGHLIGHTS
JULY 2018

OPEN SCIENCE BY DESIGN

Realizing a Vision for 21st Century Research

Openness and sharing of information are fundamental to the progress of science and to the effective functioning of the research enterprise. The advent of scientific journals in the 17th century helped power the Scientific Revolution by allowing researchers to communicate across time and space, using the technologies of that era to generate reliable knowledge more quickly and efficiently. Harnessing today's stunning, ongoing advances in information technologies, the global research enterprise and its stakeholders are moving toward a new open science ecosystem. Open science aims to ensure the free availability and usability of scholarly publications, the data that result from research, and the methodologies, including code or algorithms, that were used to generate those data.

This report provides guidance to the research enterprise and its stakeholders as they work to achieve open science.

BENEFITS AND MOTIVATION

The research enterprise has already made significant progress toward open science and is realizing a number of benefits:

- **Rigor and reliability**. New standards for data and code sharing in fields such as biomedical research and psychology are making it easier for researchers to reproduce and replicate reported work.
- **Ability to address new questions**. Open science allows researchers to bring data and perspectives from multiple fields to bear on their work, opening up new areas of inquiry and expanding the opportunities for interdisciplinary collaboration.
- **Faster and more inclusive dissemination of knowledge**. The proportion of scientific articles that are openly available is increasing, which accelerates the process of disseminating research and building on results. Open publication also allows broader, more inclusive participation in research and expands the possibilities of productive research collaboration within the United States and around the world.
- **Broader participation in research**. Large-scale projects in fields such as astronomy and ecology are utilizing open data and expanding opportunities for citizen scientists to contribute to scientific advances.
- **Effective use of resources**. Reuse of data in fields such as clinical research is facilitating the aggregation of multiple studies for meta-analysis and allows for more effective testing of new hypotheses.
- **Improved performance of research tasks**. New tools such as electronic lab notebooks enable more accurate recording of research work streams and automate various data curation tasks.

- **Open publication for public benefit**. The belief that the broader public should have access to publicly-funded research and its benefits provides an additional strong rationale for open science. In the case of publicly-funded research, the ultimate sponsor is the taxpayer. The public benefits from open science as new knowledge is utilized more rapidly to improve health, protect environmental quality, and deliver new products and services.

A number of public and private research funders have introduced policies and support systems to ensure that the results of the research they sponsor are open.

BARRIERS AND CHALLENGES

Despite the significant progress made in recent years toward creating an open science ecosystem, science today is not completely open. Several barriers and challenges remain:

- **Costs and infrastructure**. There are significant remaining cost barriers to widespread implementation of open publication and open data. New technological and institutional infrastructure within specific disciplines and across disciplines needs to be developed.
- **Structure of scholarly communications**. Most publications are still only available on a subscription basis, and some potential pathways to open publication may disrupt the current scholarly communications ecosystem, including scientific society publishers, or may disadvantage early career researchers, researchers working in the developing world, or those in institutions with fewer resources.
- **Lack of supportive culture, incentives and training**. Open practices such as preparing datasets and code for sharing and making preprints available are not generally rewarded and may even be discouraged by current incentive and reward systems. This may have the unintended consequence of causing a disadvantage to early career researchers.
- **Privacy, security, and proprietary barriers to sharing**. Sharing data, code, and other research products is becoming more common, but barriers related to ensuring patient confidentiality and the protection of national security information exist in some domains. Proprietary research also presents barriers. Ultimately, some parts of the research enterprise may not be open.
- **Disciplinary differences**. The nature of research and practices surrounding treatment of data and code differ by discipline and even within a discipline. The size of datasets and the nature of some data may prevent immediate, complete sharing. Safeguards to prevent misuse or misrepresentation of data will be needed.

Open science stands at an important inflection point. A new generation of information technology tools and services holds the potential of further revolutionizing scientific practice. The ability to automate the process of searching and analyzing linked articles and data can reveal patterns that would escape human perception, making the process of generating and testing hypotheses faster and more efficient. These tools and services will have maximum impact when used within an open science ecosystem that spans institutional, national, and disciplinary boundaries.

OPEN SCIENCE BY DESIGN

In order to frame the issues and possible actions, the committee developed the concept of open science by design, defined as a set of principles and practices that fosters openness throughout the entire research life cycle (see Figure 1).

The researcher is at the center of the concept of open science by design. At each stage of the research process, the researcher both contributes to open science and takes advantage of the open science practices of other members of the research community:

- **Provocation:** explore or mine open research resources and use open tools to network with colleagues.
- **Ideation:** develop and revise research plans and prepare to share research results and tools under FAIR (Findable-Accessible-Interoperable-Reusable) principles.
- **Knowledge generation:** collect data, conduct research using tools compatible with open sharing, and use automated workflow tools to ensure accessibility of research outputs.
- **Validation:** prepare data and tools for reproducibility and reuse and participate in replication studies.

- **Dissemination:** use appropriate licenses for sharing research outputs and report all results and supporting information (data, code, articles, etc.).
- **Preservation:** deposit research outputs in FAIR archives and ensure long-term access to research results.

The overarching principle of open science by design is that research conducted openly and transparently leads to better science. The vision of open science by design suggests that all phases of the research process provide opportunities for assessing and improving the reliability and efficacy of scientific research.

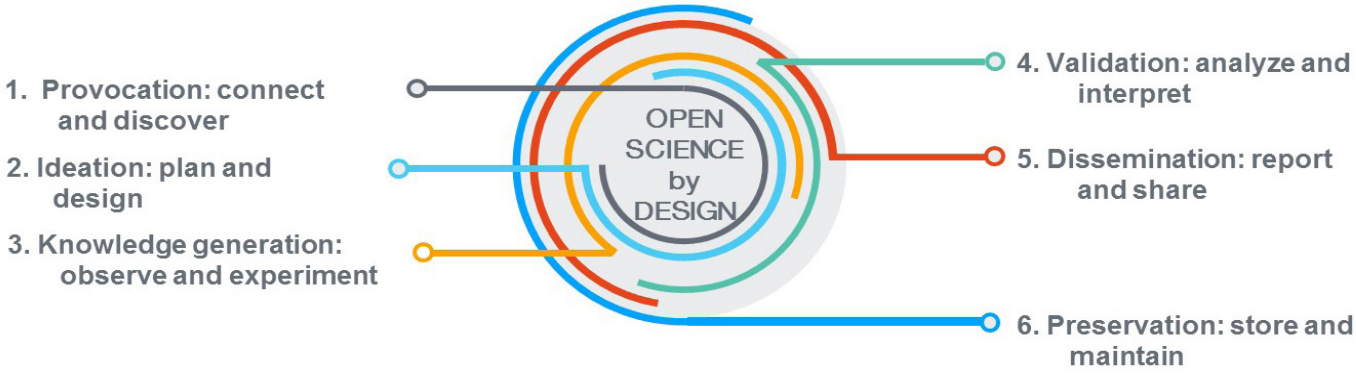

FIGURE 1 Phases of Open Science by Design in the research life cycle.
SOURCE: Committee generated.

ACCELERATING PROGRESS

Achieving open science will require persistent, coordinated actions on the part of research enterprise stakeholders. The committee's key findings, recommendations, and implementation actions are listed below.

Building a Supportive Culture

FINDING: Continued effort by stakeholders, working internationally and across disciplinary boundaries, is needed to change evaluation practices and introduce other incentives so that the cultural environment of research better supports and rewards open practices.

> **RECOMMENDATION ONE: Research institutions should work to create a culture that actively supports Open Science by Design by better rewarding and supporting researchers engaged in open science practices. Research funders should provide explicit and consistent support for practices and approaches that facilitate this shift in culture and incentives.**

Training for Open Science by Design

FINDING: There is little formal training and education in the principles and practices of open science. The emergence of data science as a recognized interdisciplinary field has highlighted the need for new educational content and approaches related to data.

> **RECOMMENDATION TWO: Research institutions and professional societies should train students and other researchers to implement open science practices effectively and should support the development of educational programs that foster Open Science by Design.**

Ensuring Long-term Preservation and Stewardship

FINDING: Developing and sustaining the infrastructure required for long-term stewardship of research products will present a continuing challenge.

> **RECOMMENDATION THREE: Research funders and research institutions should develop the policies and procedures to identify the data, code, specimens, and other research products that should be preserved for long-term public availability, and they should provide the resources necessary for the long-term preservation and stewardship of those research products.**

Facilitating Data Discovery, Reuse, and Reproducibility

FINDING: As progress toward opens science by design continues, it is important that the community adhere to the ultimate goal of achieving the availability of research products under open principles.

> **RECOMMENDATION FOUR: Funders that support the development of research archives should work to ensure that these are designed and implemented according to the FAIR data principles. Researchers should seek to ensure that their research products are made available according to the FAIR principles and state with specificity any exceptions based on legal and ethical considerations.**

Developing New Approaches to Fostering Open Science by Design

FINDING: Public and private funders have made significant contributions to fostering open science to this point. They should continue to support initiatives that accelerate progress, and evaluate and revise their policies as needed.

> **RECOMMENDATION FIVE: The research community should work together to realize Open Science by Design to advance science and help science better serve the needs of society.**

COMMITTEE ON TOWARD AN OPEN SCIENCE ENTERPRISE

Alexa T. McCray (*Chair*), Harvard Medical School; **Francine Berman**, Rensselaer Polytechnic Institute; **Michael Carroll**, American University Washington College of Law; **Donna Ginther**, University of Kansas; **Robert Miller**, LYRASIS; **Peter Schiffer**, Yale University; **Edward Seidel**, University of Illinois at Urbana-Champaign; **Alex Szalay**, The Johns Hopkins University; **Lisa Tauxe**, University of California, San Diego; **Heng Xu**, The Pennsylvania State University. Staff: **Tom Arrison**, Program Director, Policy and Global Affairs Division (from November 2017); **Emi Kameyama**, Associate Program Officer, Board on Research Data and Information; **George Strawn**, Director, Board on Research Data and Information; **Ester Sztein**, Deputy Director, Board on Research Data and Information; **Nicole Lehmer**, Senior Program Assistant, Board on Research Data and Information; **Alan Anderson**, Consultant; and **Christine Liu**, Senior Program Officer (until October 2017).

For More Information . . . This Consensus Study Report Highlights was prepared by the Board on Research Data and Information based on the Consensus Study Report *Open Access by Design: Realizing a Vision for 21st Century Research* (2018). The study was sponsored by the Laura and John Arnold Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project. Copies of the Consensus Study Report are available from the National Academies Press, (800) 624-6242; http://www.nap.edu or via the Board on Research Data and Information web page at http://www.nationalacademies.org/PGA/BRDI.

Board on Research Data and Information
Policy and Global Affairs

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.
www.national-academies.org

Copyright 2018 by the National Academy of Sciences. All rights reserved.