



BERKELEY LAB

Bringing Science Solutions to the World



U.S. DEPARTMENT OF
ENERGY

Office of Science

Data publishing and ethical challenges addressing datasets

Joerg Heber

Research Compliance Officer

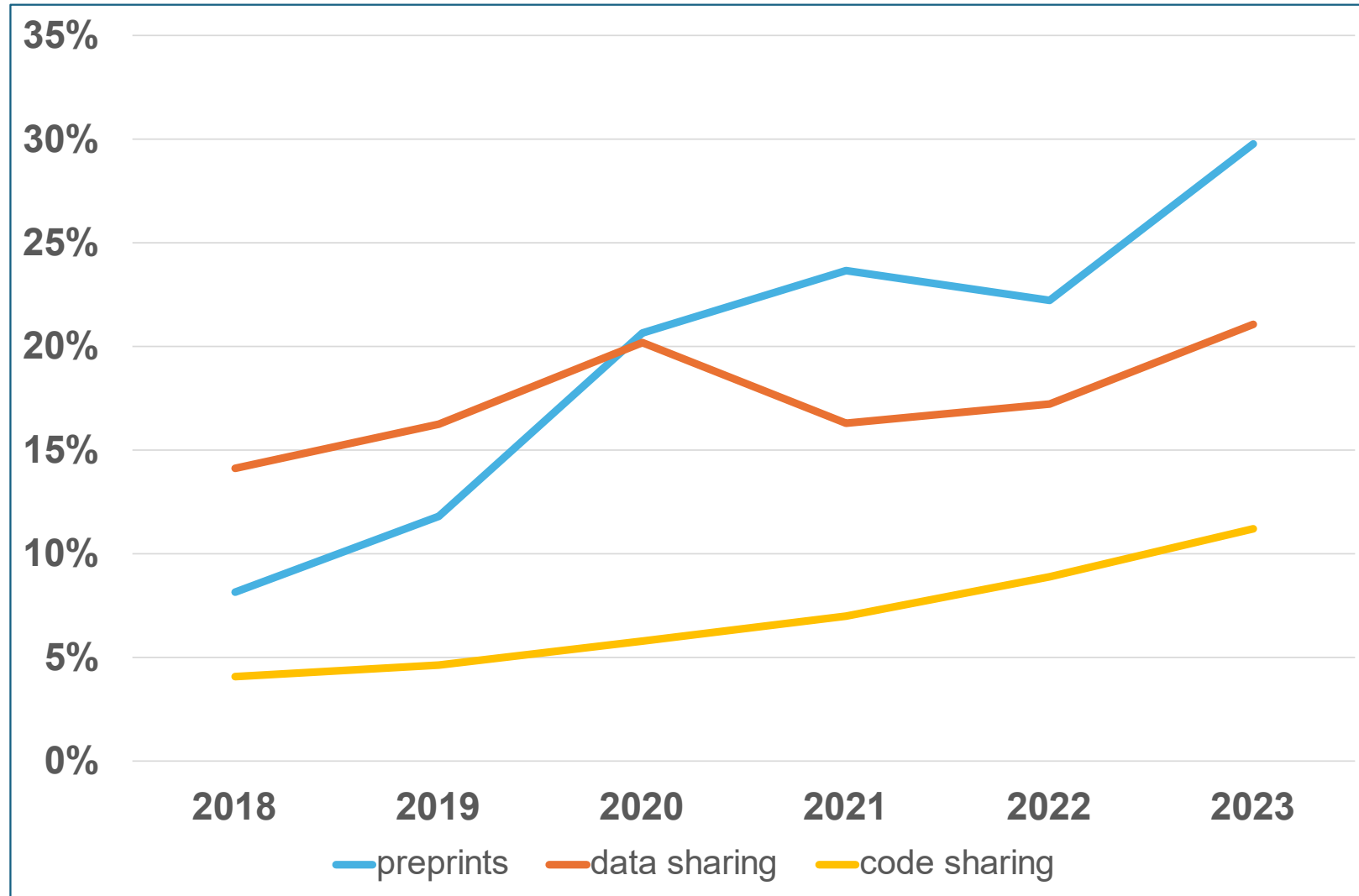
Lawrence Berkeley National Laboratory

Open Science & Data Sharing

Open Science: Beyond the Published Article



Increase in open research practice



Is there a benefit to data sharing?

The citation advantage of linking publications to research data

Giovanni Colavizza, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, Barbara McGillivray 

Published: April 22, 2020 • <https://doi.org/10.1371/journal.pone.0230416>

Computer Science > Digital Libraries

[Submitted on 24 Apr 2024]

An analysis of the effects of sharing research data, code, and preprints on citations

[Giovanni Colavizza](#), [Lauren Cadwallader](#), [Marcel LaFlamme](#), [Grégory Dozot](#), [Stéphane Lecorney](#), [Daniel Rappo](#), [Iain Hrynaszkiewicz](#)

<https://arxiv.org/abs/2404.16171>

Large-scale studies find citation advantages for data sharing of ~ 4-25% and for posting preprints of ~20%.

OSTP Guidance

AUGUST 25, 2022

OSTP Issues Guidance to Make Federally Funded Research Freely Available Without Delay

U.S. policy guidance to make the results of taxpayer-supported research immediately available to the American public at no cost.

DOE Public Access Plan

- All DOE-funded research are subject to Data Management and Sharing Plans (DMSP)
- DMSP will address validation and replication of results, timely and equitable access, data repository selection, data management resources, and data sharing limitations.
- Reviewers of Office of Science research proposals evaluate DMSP



ENERGY.GOV

<https://www.energy.gov/doe-public-access-plan>

Data Sharing at Berkeley Lab

- Openness in science part of our core values
- Global impact of the Lab:
 - \$1.02 billion budget
 - 1,800 researchers
 - 14,000 researchers worldwide use our research facilities
- Research Data Policy since 2022:
 - Mandates DMSP
 - Encourages sharing of research data via FAIR Principles
 - Data ownership and processes when researchers leave LBNL
 - Establishes RIO as point of contact regarding data sharing concerns



Ethical Challenges Addressing Data Sets



Global Working Group on the Handling of Ethical Concerns Relating to the Publication of Research Data

Working group that brought together publishers, repositories and institutions

Outcome Documents:

<https://force11.org/post/recommendations-for-the-handling-of-ethical-concerns-relating-to-the-publication-of-research-data/>

<https://zenodo.org/records/5391293>

Who owns the data?

- Data is typically owned by the institution
- PI are the stewards of that data, with authority to make decisions
 - Institutions are part of the dialogue around research data
 - Institutions also have responsibilities for data

Data Issues: Authorship & Contribution Conflicts

- Journals have established processes for research integrity & conflicts
 - What about data publishers?
 - Author lists of data sets can be different than author lists on related journal article
 - Datasets can have many more contributors (e.g. citizen science)
 - Author lists should be changeable (e.g. name changes, dispute resolutions)
-
- Coordination needed if conflicts impact different research outputs
 - No formal dispute process may exist outside of institutions
 - Do institutions have processes in place?

Data Issues: Legal & Regulatory Concerns

- What if there are legal or regulatory concerns about data sets?
 - Copyright concerns (including generative AI!)
 - Local legal restrictions
 - Privacy and ethics concerns (participants in human subjects research)
 - Sensitive subjects (e.g. hate speech as research topic)
 - Problematic data origin (e.g. data arising from surveillance practices)
 - Sensitive data (location of endangered species)
 - Dual Use Research of Concern, research security
- Data publishers responsible to address concerns
- Needs involvement of authors, institutions if necessary

Data Issues: Research Ethics and Integrity Concerns

- Fabrication, Falsification and Plagiarism (naturally) impact research data sets
- Datasets usually relate to investigations by institutions, funders, journals
- Institutions need to consider data outputs in addition to journal articles in correcting the scholarly record
- Processes for data removals/replacements, or retraction?
- Who has responsibility?
- Consider where you deposit your data!



Questions for institutions

Open Research and Research Ethics

- There are an increased number of stakeholders for research outputs
- Data sharing can lead to inadvertent releases of sensitive data
- Access controls can mitigate privacy and other concerns
- Research ethics and integrity issues need to be addressed for all research outputs

The role of institutions

- Ensure all research outputs are considered when addressing ethics concerns
- Who has responsibility to resolve concerns arising from data sets?
- How are conflicts resolved? Do authorship guidelines cover data sets?
- Is it clear to all parties whom to contact at institutions?
- Do researchers have sufficient guidance?

LBL Research Data Policy – Contacts Section

- Questions around the use, management, and sharing of research data should be directed to the relevant PI or to the applicable research division, data repository, or facility.
- Questions around licensing or patenting of research data should be directed to the [Intellectual Property Office \(IPO\)](#) at ipo@lbl.gov.
- Questions about the implementation of this policy should be directed to the [Research Compliance Office \(RCO\)](#) at rio@lbl.gov.
- Concerns or disputes over the control, use, and integrity of research data or violations of this policy should be directed to the Research Integrity Officer (RIO) at rio@lbl.gov. The RIO may discuss concerns informally, which may include discussing them anonymously and/or hypothetically. If the circumstances described by an individual do not suggest research policy violations, the RIO will refer the matter to other offices or officials with responsibility for resolving a concern.
- Questions about applicable export control regulations should be directed to the [Export Compliance Office \(ECO\)](#) at exportcontrol@lbl.gov.
- For security concerns or in cases of accidental disclosure of personally identifiable information or protected health information, contact it-policy@lbl.gov.
- Questions about records retention should be directed to the [Records Management Office \(RMO\)](#) at records@lbl.gov

THANK YOU!

jheber@lbl.gov