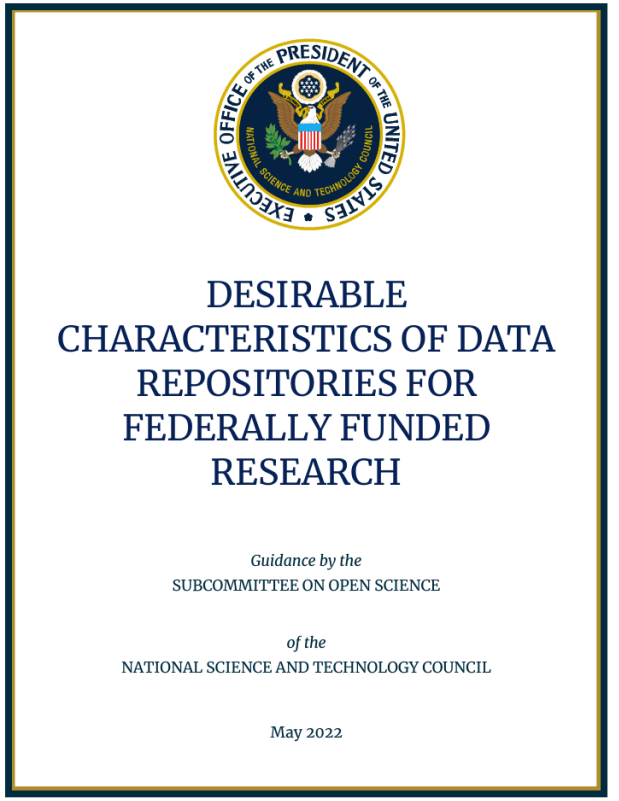
FAIR Data Repositories: Expectations, Obligations, and Expenses

Robert Hanisch
Director, Office of Data and Informatics
NIST

National Institute of Standards and Technology
U.S. Department of Commerce

Expectations

- Holdren memo, OSTP, February 2013
- FAIR principles, March 2016
- “Desirable Characteristics,” May 2022, NSTC Subcommittee on Open Science
- Nelson memo, OSTP, August 2022
  - “Scientific data underlying peer-reviewed scholarly publications resulting from federally funded research should be made freely available and publicly accessible by default at the time of publication”
- Units

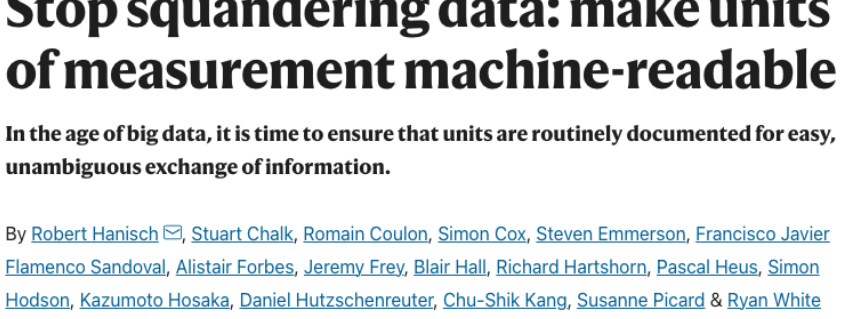


DESIRABLE CHARACTERISTICS OF DATA REPOSITORIES FOR FEDERALLY FUNDED RESEARCH

Guidance by the
SUBCOMMITTEE ON OPEN SCIENCE

of the
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

May 2022



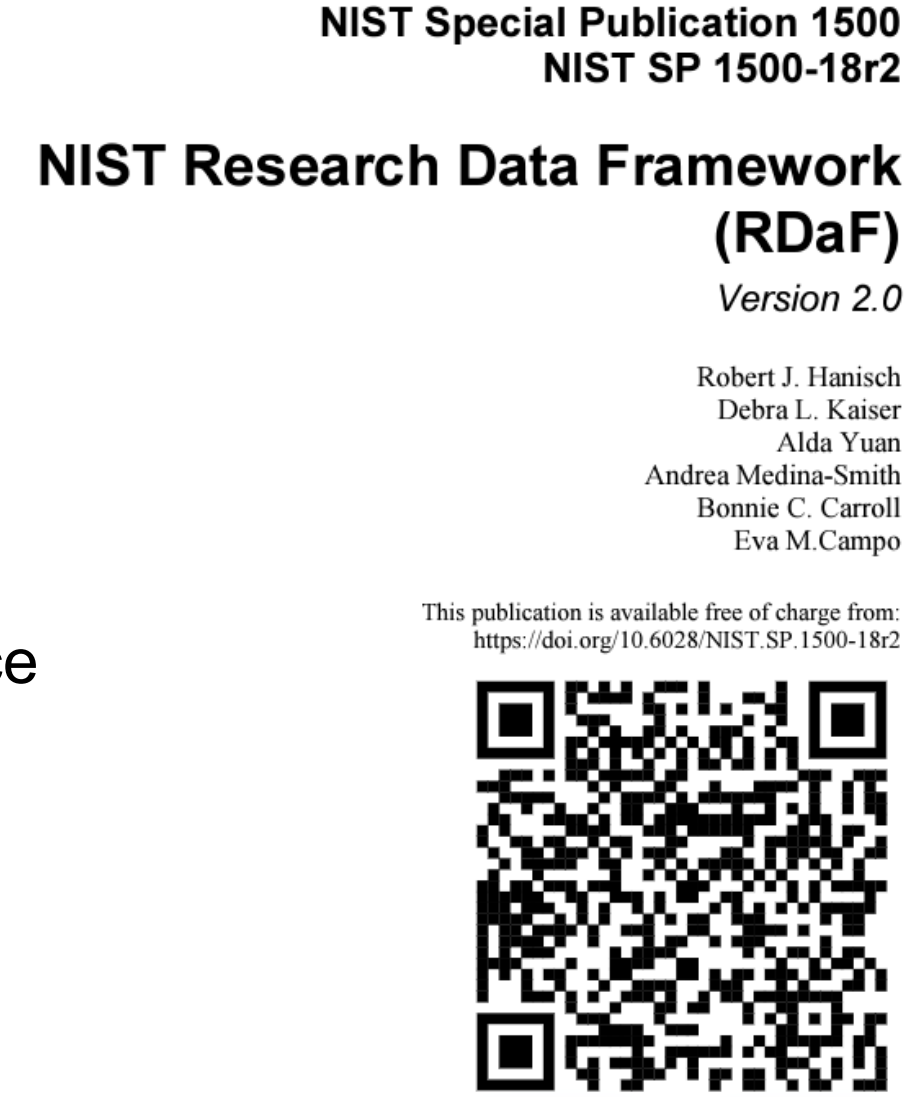
Stop squandering data: make units of measurement machine-readable

In the age of big data, it is time to ensure that units are routinely documented for easy, unambiguous exchange of information.

By Robert Hanisch, Stuart Chalk, Romain Coulon, Simon Cox, Steven Emmerson, Francisco Javier Flamenco Sandoval, Alistair Forbes, Jeremy Frey, Blair Hall, Richard Hartshorn, Pascal Heus, Simon Hodson, Kazumoto Hosaka, Daniel Hutzschenreuter, Chu-Shik Kang, Susanne Picard & Ryan White

Obligations

- Data management plans
- Data born FAIR
  - Laboratory Information Management Systems
  - Data models
  - Metadata standards
  - Open (non-proprietary) data formats
  - FAIR Digital Objects (FDOs)
- Units of Measurement Interoperability Service
  - Fundamental constants
- Research data management, principles and practices: NIST RDaF



NIST Special Publication 1500
NIST SP 1500-18r2

NIST Research Data Framework (RDaF)

Version 2.0

Robert J. Hanisch
Debra L. Kaiser
Alda Yuan
Andrea Medina-Smith
Bonnie C. Carroll
Eva M.Campo

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.1500-18r2

Obligations

- Federal agencies supported $54B in university-based research (2022)*
- ~10% of this supports research publication costs (APCs, subscriptions)
  - Elsevier income: €2.91B in 2022
  - STM estimates journal market value of $10B
- Data are an essential component of the research record
  - Reproducibility, reliability, robustness, rigor; transparency, trust
  - Open source software
- Quality data are the fodder for AI
- Funders should be *compelled* to set aside long-term support for data repositories

* https://ncses.nsf.gov/pubs/nsf24307

Expenses

- We often hear that data curation and preservation is too expensive; not true!
- In astronomy, for example, DC&P varies between 1 and 10% of a facility's annual operating budget
  - And well-curated data gets reused; large multiplier for ROI
- At NIST, we built and operate our Public Data Repository for ~0.1% of our annual research budget
- An annual investment of 2-3% of the federal research budget would solve the problem!
  - Positive impact on research
  - Data can be reused, repurposed for a fraction of the cost of the original acquisition