

NIST Data Management and Sharing and the Research Data Framework (RDaF)

Dr. Robert Hanisch

Director, Office of Data and Informatics (ODI)

Material Measurement Laboratory

US National Institute of Standards and Technology



About NIST

- The National Institute of Standards and Technology is a federal agency under the US Department of Commerce
 - Known as the National Bureau of Standards until 1988, originally founded in 1901
- Non-regulatory
- State of the art in measurement science and technology
- US National Metrology Institute, amongst network of 103 NMIs globally organized under the Bureau International des Poids et Mesures (BIPM, or International Bureau of Weights and Measures), Paris
- ~5,000 staff at NIST (Gaithersburg, Maryland headquarters; Boulder, Colorado; Charleston, South Carolina; Brookhaven National Laboratory)
- 6 major research laboratories
 - Material Measurement Laboratory
 - Office of Data and Informatics (15 people)



Science Data Portal and Public Data Repository

- Modern website for search and discovery of NIST public data sets
 - <https://data.nist.gov>
- Developed and operated by ODI for NIST
 - Front end to the NIST Public Data Repository
 - Implements the NIST taxonomy for research domains
- Open-source code base – hosted on [github/USNISTGOV](https://github.com/USNISTGOV)

NIST SCIENCE DATA PORTAL 1.4.1

Key Datasets ▾ Standard Reference Data (SRDs) Developer ▾ About ▾ Find Papers ▾ Queries 0

NIST Data Discovery

Explore data, tools, and resources for Science, Engineering, Technology and more

Q Gallium ALL RESEARCH Search Advanced Search

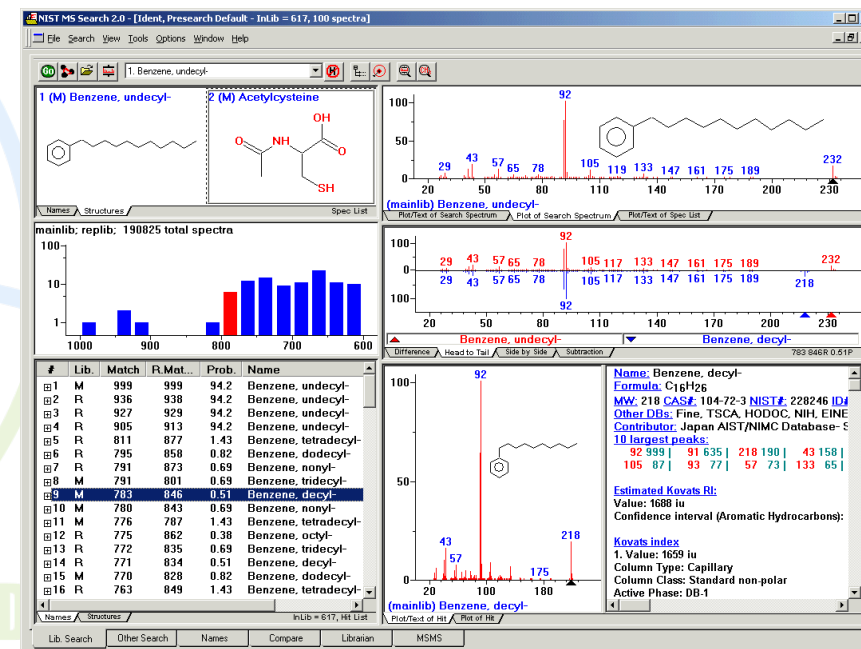
Examples: "Kinetics database" Gallium "SRD 101" XPDB Interatomic Potentials

FEATURED DATA DOMAINS

| | | | |
|------------------------|----------------------------|-------------------------|-----------|
| INFORMATION TECHNOLOGY | MATHEMATICS AND STATISTICS | MANUFACTURING | FORENSICS |
| MATERIALS | PHYSICS AND NEUTRON | ADVANCED COMMUNICATIONS | CHEMISTRY |

NIST Standard Reference Data

- Most highly vetted data products of NIST
 - SRD Act of 1968
- 65 databases, free and subscription based
- 6,000 units sold/year as downloads and agreements including royalties on instrument sales
- Online SRD Metrics
 - 2M views a month webbook.nist.gov
 - 300K views a month [XPS - NIST X-Ray Photoelectron Spectroscopy Database](https://www.nist.gov/xps)



Stephen E. Stein (2014), NIST/EPA/NIH Mass Spectral Library with Search Program - SRD 1a, National Institute of Standards and Technology, <https://doi.org/10.18434/T4H594> (Accessed 2020-09-08)

What is a Research Data Framework?

- A map of the research data space: who, what, where, why, when?
- A dynamic guide for the various stakeholders in research data to understand best practices for research data management and dissemination
- A resource for understanding costs, benefits, and risks associated with research data management
- A consensus document based on inputs and conversations amongst the stakeholders in research data

Why a Research Data Framework?

- Research data ecosystem is very complex!
 - Lots of players, various funding models and sustainability plans
 - How long should data be kept?
 - How should data quality be assessed?
 - How do we measure the value of research data?

NIST

PROJECTS/PROGRAMS

Research Data Framework (RDaF)

Summary

In the past decade, research data have become widely recognized as a critical national and global resource, and the risks of losing or mismanaging research data can have severe economic and social consequences. The proliferation of artificial intelligence approaches in all fields has created a huge demand for trustworthy research data in both the natural (e.g., chemistry) and social (e.g., economics) sciences. To address these issues, NIST initiated a new, multi-stakeholder project in fall 2019 entitled the Research Data Framework (RDaF). The RDaF will provide the stakeholder community with a structured approach to develop a customizable strategy for various roles in the research data management ecosystem.

[Download the Preliminary RDaF document \(PDF\).](#)

DESCRIPTION

What is a Research Data Framework?

- A map of the research data space: who, what, where, why, when?
- A dynamic guide for the various stakeholders in research data to understand best practices for research data management and dissemination
- A resource for understanding costs, benefits, and risks associated with research data management
- A consensus document based on inputs and conversations amongst the stakeholders in research data
- A tool that may be used to change the research data management culture in an organization



Stakeholders

- Government agencies
- National laboratories
- Universities and research libraries
- Data repositories
- Scholarly publishers
- Professional societies
- National and international collaboration organizations (e.g., CENDI, BRDI, CODATA, RDA, WDS, GO-FAIR)
- Standards bodies
- Funders (public and private)
- Industry and the private sector
- Researchers
- General public



Why a Research Data Framework?

- Leverage research data to address global challenges



United Nations Sustainable Development Goals (SDGs)

RDaF Benefits

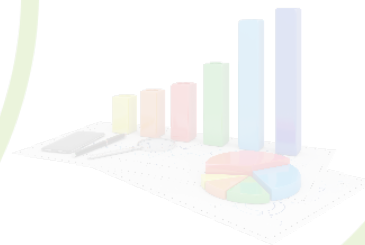
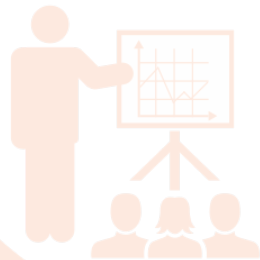
- **Increase research integrity** with quality data and improved transparency of the research process
- **Reduce costs and maximize efficiency** by establishing best practices for data management
- **Guide risk management and reduction** through assessment of risk positions and roadmaps for improvement
- **Increase scientific discovery and innovation** with the FAIR principles (Findable, Accessible, Interoperable, Reusable) for better utilization of data

National and International Need

- Data is proliferating at an exponential rate
- Data management is complex and confusing
- Mismanaged data has dire social and economic consequences, including loss of global leadership in critical technical fields
- The U.S. needs a coordinated effort to establish a research data infrastructure, but research data are global in nature so international collaboration / coordination is necessary
- NIST is well-positioned to lead the project; our business is consensus building through being a neutral convener of diverse communities

Process

- Pilot program to provide an overall guide to the actors and stakeholders in the research data space
- NIST Cybersecurity Framework is the model
- Community consensus, not NIST imposition
- If I am a _____, then I need to know _____.
- Initial scoping workshop held in December 2019 at NIST
 - 50 invited participants representing stakeholders, both US and international



RDaF Publication

Framework Core

Functions (Six Research Data Lifecycle Stages)

- **Categories** (topics)
- **Subcategories** (sub-topics)
- Informative References

1. Introduction
 - 1.1. Motivation
 - 1.2. Origin of the Framework
 - 1.3. What is the RDaF?
 - 1.4. Legal and Institutional Drivers
 - 1.5. Value Proposition
 - 1.6. Risk Management
2. Development of the Preliminary RDaF
 - 2.1. Initial Scoping Study
 - 2.2. Stakeholder Scoping Workshop
 - 2.3. Interim Studies and Reports
 - 2.4. Drafting the Preliminary RDaF
3. Description of the Preliminary RDaF
 - 3.1. Relationship to Other NIST Frameworks
 - 3.2. Framework Core
 - 3.3. Informative References
 - 3.4. Framework Profiles
 - 3.5. Framework Implementation Tiers
4. Next Steps

NIST Special Publication 1500-18

Research Data Framework (RDaF): Motivation, Development, and A Preliminary Framework Core

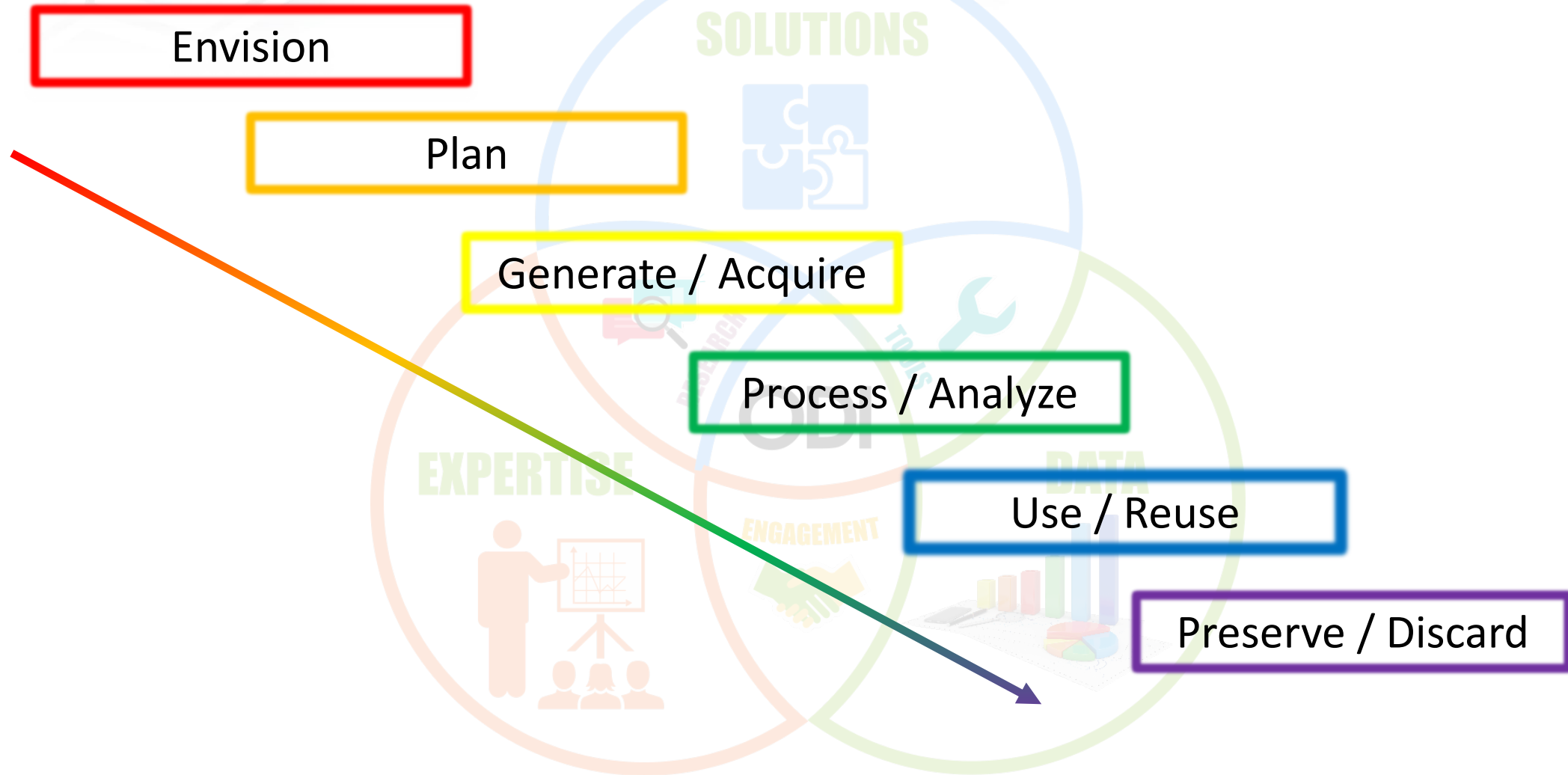
Robert J. Hanisch
Debra L. Kaiser
Bonnie C. Carroll

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

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

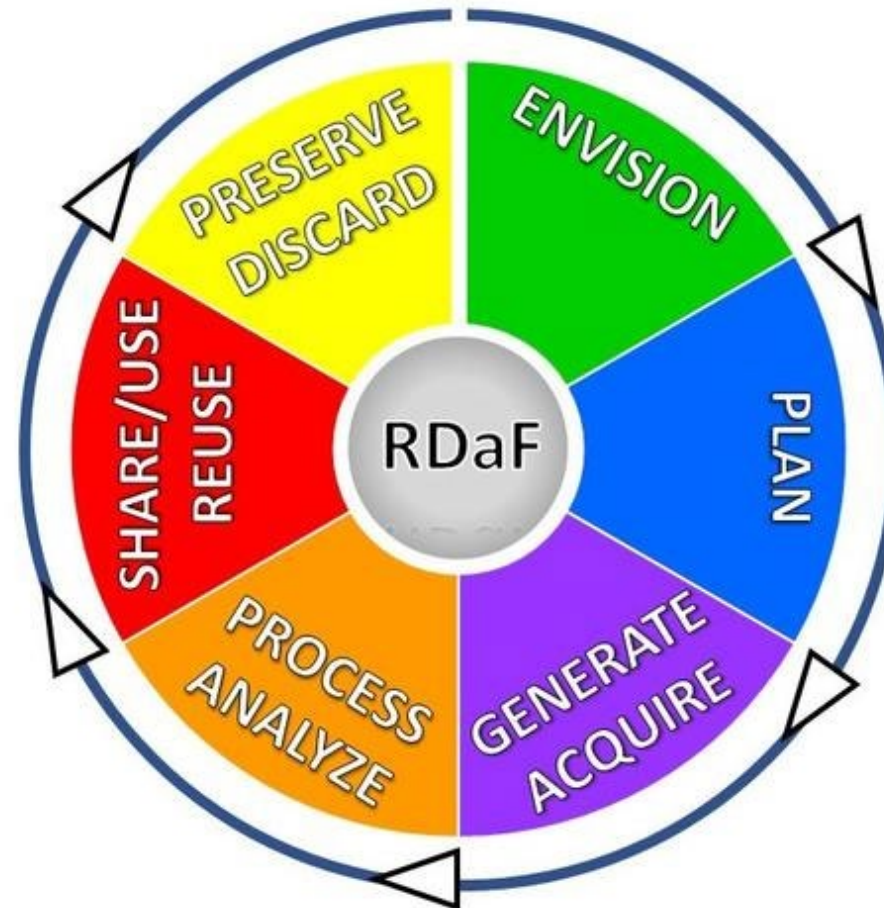
<https://doi.org/10.6028/NIST.SP.1500-18>

RDaF Structure Based on “Functions”



RDaF Lifecycle Stages

- 1) **Envision** - chart a high-level course of action to achieve desired organizational goals
- 2) **Plan** - preparation in an organization for effective research data management
- 3) **Generate/Acquire** - generation of raw research data, both experimentally and computationally, within an organization, and the collection or acquisition of research data produced outside of an organization



- 4) **Process/Analyze** - actions performed on generated or acquired research data to yield processed research data
- 5) **Share/Use/Reuse** - outlines how raw and processed research data are disseminated, used, and reused within an organization and any constraints or encouragements to use/reuse
- 6) **Preserve/Discard** - end-of-use and end-of-life provisions for research data in an organization and includes records management, archiving, and safe disposal

Envision

Function

1) Envision

Category

*Data governance—
strategic/qualitative*

*Data governance—
legal and regulatory
compliance*

*Community
engagement*

*Data culture and
reward structure*

Subcategory

Data vision, data policy
Data management organization
Data quality, data stewardship

Data privacy and ethics
Data inventory
Risk assessment and management

Communication, interactions
Cross-domain

FAIR principles
Value of data and data professionals
Roles and responsibilities
Incentives for sharing and re-use

Status

- Held two Opening Plenary Workshops as pilots: materials science (12/10/21) and research universities/libraries/scholarly publishers (10/31/21)
- Strong interest and engaged participation from stakeholder groups, e.g.,
 - Government departments/agencies (DOD, DOE, NASA, NIH, NSF)
 - Professional societies and trade organizations (AGU, ACerS, APS, ASM International, MRS, AAU, APLU, ARL)
 - Scholarly publishing community (AAAS/Science, AIP Publishing, Elsevier, Springer Nature)
- Generated RDaF version 1.1 with input from the workshop participants

<https://www.nist.gov/programs-projects/research-data-framework-rdaf>

Next Steps

- Hold 15 Stakeholder Workshops, each focused on a specific stakeholder role, e.g., Senior Executives, Researchers, Scholarly Publishers
- Two-hour-long workshops with 10-12 participants, in the June to mid-July timeframe
- NIST team will develop preliminary “**Profiles**” for each stakeholder role—essentially a checklist of those categories and subcategories across the research data lifecycle that are most relevant for that role.
- Participants will be asked to provide input on the following:
 - The preliminary Profile for their specific job role
 - The entire RDaF Core version 1.1 (all categories and subcategories)
 - What elements of the Framework do you influence? What elements influence you? Where in the Framework are your primary responsibilities represented?
- After the Stakeholder Workshops, the NIST team will develop RDaF version 1.2 and finalize a set of Profiles for representative job roles

<https://www.nist.gov/programs-projects/research-data-framework-rdaf>

Partial Preliminary Profile for a Senior Executive

| FUNCTION (Data Lifecycle Stage) | CATEGORY | SUBCATEGORY | Relevancy (0 to 3)* |
|---|---|--|---------------------|
| ENVISION Review of the overall strategies and drivers of an organization's research data program. | Data Governance— Strategic/Qualitative | Identification of Goals and Roles | 2 |
| | | Data vision and/or data policy | 3 |
| | | Data management organization | 2 |
| | | Data quality | 1 |
| | | Data stewardship | 1 |
| | Data Governance— Legal and Regulatory Compliance | Data privacy | 3 |
| | | Safety and security assurance | 3 |
| | | Data inventory | 0 |
| | | Risk mitigation and management | 1 |
| | Data Culture and Reward Structure | Roles and responsibilities | 2 |
| | | Value of data to organization and leadership | 3 |
| | | Disincentives for data sharing | 0 |
| | Resources—Allocation and Sustainability | Sources of funding | 3 |
| | | Long-term funding | 3 |

*scale: 3 denotes the greatest relevance and 0 is not relevant

RDaF Steering Committee (SC)



Bonnie Carroll
CODATA, SC Chair



Laura Biven
NIH



Cate Brinson
Duke Univ



Martin Halbert
NSF



Hilary Hanahoe
RDA



Heather Joseph
SPARC



Mark Leggott
Research Data Canada



Barend Mons
CODATA, GO-FAIR



Sarah Nusser
Iowa State U, UVA



Beth Plale
Indiana Univ



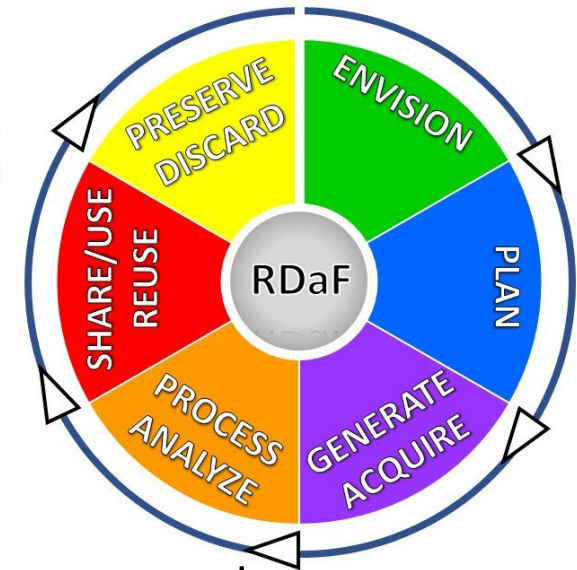
Carly Strasser
Zuckerberg Inst.



Anita de Waard
Elsevier

RDaF Summary

- Successful in building community interest and engagement
 - Diverse stakeholders
 - National and international
- Challenges
 - Resources
 - Timeliness: the research data ecosystem is changing rapidly. How to keep pace and assure ongoing updates?
 - Controlling scope and scale
- Next steps
 - Hold a series of Stakeholder Workshops focused on specific job roles to obtain feedback on RDaF V1.1 and to develop role-based profiles, i.e., checklists of key topics for a specific job role; integrate findings into V1.2
 - Solicit community input through request for information, Federal Register
 - Hold plenary workshop to validate updated Framework
 - Publish V2.0



Contacts

SOLUTIONS

Robert Hanisch

robert.hanisch@nist.gov

<https://nist.gov/people/robert-hanisch>

ODI

Debra Kaiser, debra.kaiser@nist.gov

Bonnie Carroll, bcarrolltn@gmail.com