

# NIST Research Data Framework (RDaF)

Dr. Robert Hanisch

Director, Office of Data and Informatics (ODI)

Material Measurement Laboratory

US National Institute of Standards and Technology



# 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

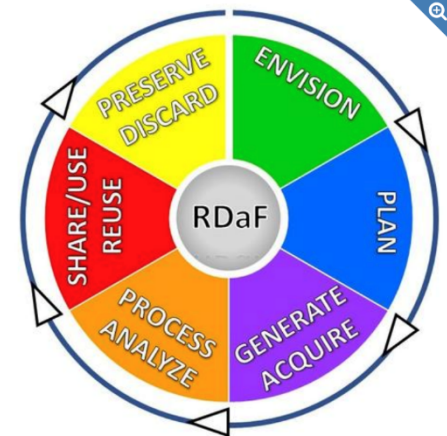
In the past decade, research data have become widely recognized as a critical national and global resource, and the risks of losing or mismanaging research data can have severe economic and social consequences. The proliferation of artificial intelligence approaches in all fields has created a huge demand 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 service/infrastructure providers, e.g., data repositories
- Scholarly publishers
- Professional societies and trade organizations
- National and international collaboration organizations (e.g., CENDI, BRDI, CODATA, RDA, WDS, GO-FAIR)
- Non-profit organizations and NGOs
- Standards bodies
- Funders (public and private)
- Industry and the private sector
- Researchers
- General public





# 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



## Research Data Framework

Robert Hanisch  
Director, Office of Data  
and Informatics  
Material Measurement  
Laboratory  
National Institute of  
Standards and  
Technology



Bonnie Carroll  
Founder & CDO,  
Information  
International  
Associates  
Secretary General,  
CODATA



# 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



# Workshop Summary

- **Status:** Confirmed support by government agencies, academic organizations, private sector companies, not-for-profit organizations, and international stakeholders.
- **Next Steps:** Management commitment to complete the scoping, pilot testing, and community building for the Framework.
  - Proposed pilots
    - Materials science
    - Universities and research libraries (AAU, APLU, ARL)

---

*Will need cooperation across government to move fully forward with the Framework.*

# 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

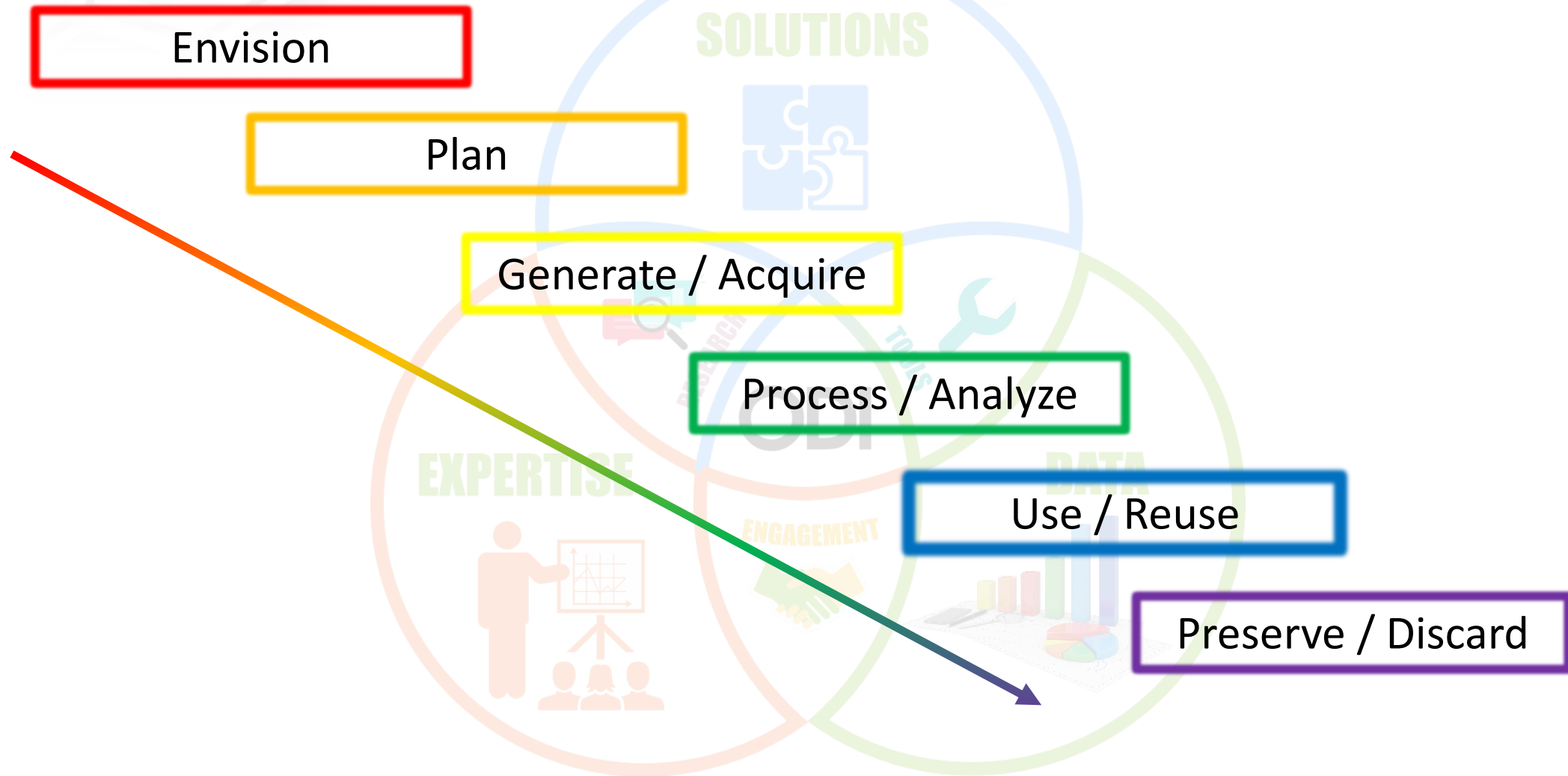
Robert J. Hanisch  
Debra L. Kaiser  
Bonnie C. Carroll

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>

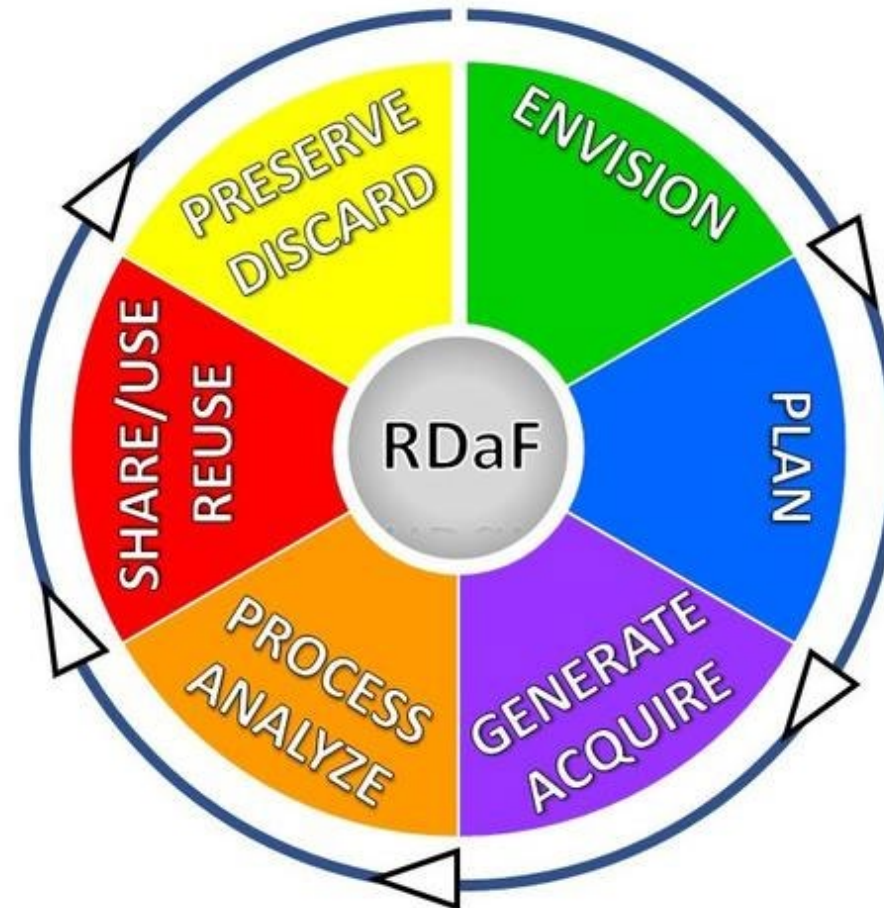
# RDaF Structure Based on “Functions”





# 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

# RDaF Structure

## 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

# RDaF Structure

## Function

## Category

## Subcategory

### 2) Plan

*Financial aspects  
of planning*

Cost-benefit analysis  
Costs by data lifecycle stage  
Funding models, funding sources

*Data objects*

Data (quantitative and qualitative)  
Software, models  
Observations, surveys

*Data management  
planning*

DMPs (intent, update)  
Metadata, analysis tools, workflows

*Data architecture*

LIMS, configuration management  
Cloud storage and computing



# RDaF Structure

## Function

### 3) Generate / Acquire

## Category

*Types of raw data*

*Sources of data*

*Acquired data*

*FAIR principles*

*Community-based  
standards*

## Subcategory

Measurements, text files  
Images, audio recordings, photos/videos

In-house and remote generation  
Generation at a user facility

Collaborators, repositories  
Extraction from the literature

Data born FAIR; data made FAIR  
Data not FAIR, e.g., legacy

Data format and file structure  
General vs. domain-specific

# RDaF Structure

## Function

### 4) Process / Analyze

## Category

*Provenance*

*Scientific workflow  
processes*

*Software tools*

*Metadata*

## Subcategory

Origin, version, time-stamp  
Data copied or derived from other data

Electronic or paper laboratory notebooks  
Containerization

Commercial or custom software  
Versioning, documentation  
Resilience, adaptability, maintenance

Responsible parties  
Linked data structure  
Persistent identification

# RDaF Structure

## Function

## Category

## Subcategory

### 5) Use / Reuse

*Legal and licenses*

Ownership, IP, rights and restrictions  
Agreements, permissions  
Citation expectations

*Data access*

Internal, external  
APIs  
Downloads vs. visiting

*New applications  
and analyses*

AI/ML  
Repurposing

*Data attribution*

Citation metrics and impact  
Provenance



# RDaF Structure

## Function

6) Preserve /  
Discard

## Category

*Sustainability*

## Subcategory

Longevity requirements  
Who pays?  
Orphan data sets

*Preservation*

Media and media migration  
Back-up  
Repositories (domain, institutional, general)  
File integrity, data recovery

*Retention and disposition*

Technical and administrative/policy decisions  
End-of-life (dark archives, deaccession, tombstones)

# Status

- Briefed OSTP Subcommittee on Open Science (01/27/22) and the Materials Research Data Alliance (MaRDA, 06/03/21)
- 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 about 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)*
<b>ENVISION</b> Review of the overall strategies and drivers of an organization's research data program.	<b>Data Governance—Strategic/Qualitative</b>	Identification of Goals and Roles	2
		Data vision and/or data policy	3
		Data management organization	2
		Data quality	1
		Data stewardship	1
	<b>Data Governance—Legal and Regulatory Compliance</b>	Data privacy	2
		Safety and security assurance	3
		Data inventory	0
		Risk mitigation and management	1
	<b>Data Culture and Reward Structure</b>	Roles and responsibilities	2
		Value of data to organization and leadership	3
		Disincentives for data sharing	0
	<b>Resources—Allocation and Sustainability</b>	Sources of funding	3
		Long-term funding	3

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

# Partial Preliminary Profile for a Data Publisher

<b>FUNCTION (Data Lifecycle Stage)</b>	<b>CATEGORY</b>	<b>SUBCATEGORY</b>	<b>RELEVANCY (0 to 3)*</b>
<b>SHARE/USE/REUSE</b> How research data are disseminated, used, and reused within and outside an organization.	<b>Data Publishing</b>	Repositories	2
		Referencing data/digital objects from journal articles	3
	<b>Internal and External Data Access</b>	Internal access e.g., data generator	0
		External access	0
		Programmatic access aka through Smart API	2
		Data access vs. data visiting	1
		Economic constraints	0
	<b>Legal and Licenses</b>	Ownership of data	3
		Constraints and encouragement for data use	2
		Intellectual property rights/restrictions	3
		Usage agreements/terms/licenses and required permissions	3
		Terms of Service	0
		Data sharing agreements and licensing	0
		Data citation	3

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

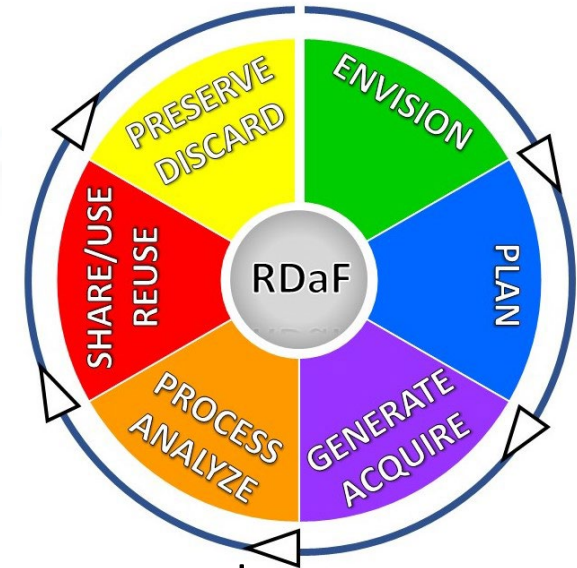
# Example of Profile Development

Research Data Lifecycle Stage	Category	Subcategory	Roles		
			CDO	Researcher	Librarian
<b>ENVISION</b> Review of the overall strategies and drivers of an organization's research data program.	<b>Data Governance: Strategic/Qualitative</b>	Data vision and/or data policy	X		
		Data management organization	X		
		Data quality	X	X	X
		Value of data	X	X	X
		Data management value proposition	X		

CDO: Chief Data Officer

# 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
- Strategy for moving forward
  - Hold a series of Stakeholder Workshops focused on specific job roles to obtain feedback on RDaF version 1.1 and to develop role-based profiles, i.e., checklists of key topics for a specific job role
  - Collaborate with other federal agencies, professional societies, scholarly publishing community, etc., to garner the necessary resources and take advantage of work in progress





# Contacts

## SOLUTIONS

Robert Hanisch

[robert.hanisch@nist.gov](mailto:robert.hanisch@nist.gov)

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

## ODI

Debra Kaiser, [debra.kaiser@nist.gov](mailto:debra.kaiser@nist.gov)

Bonnie Carroll, [bcarrolltn@gmail.com](mailto:bcarrolltn@gmail.com)

# NIST Frameworks

Framework for Improving Critical Infrastructure Cybersecurity:

<https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf>

NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, September 6, 2019 (Preliminary Draft)

[https://www.nist.gov/system/files/documents/2019/09/09/nist\\_privacy\\_framework\\_preliminary\\_draft.pdf](https://www.nist.gov/system/files/documents/2019/09/09/nist_privacy_framework_preliminary_draft.pdf)

NIST Big Data Interoperability Framework: Volume 1, Definitions October 2019 Version 3

<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf>