

# 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



# About NIST and ODI

- 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)



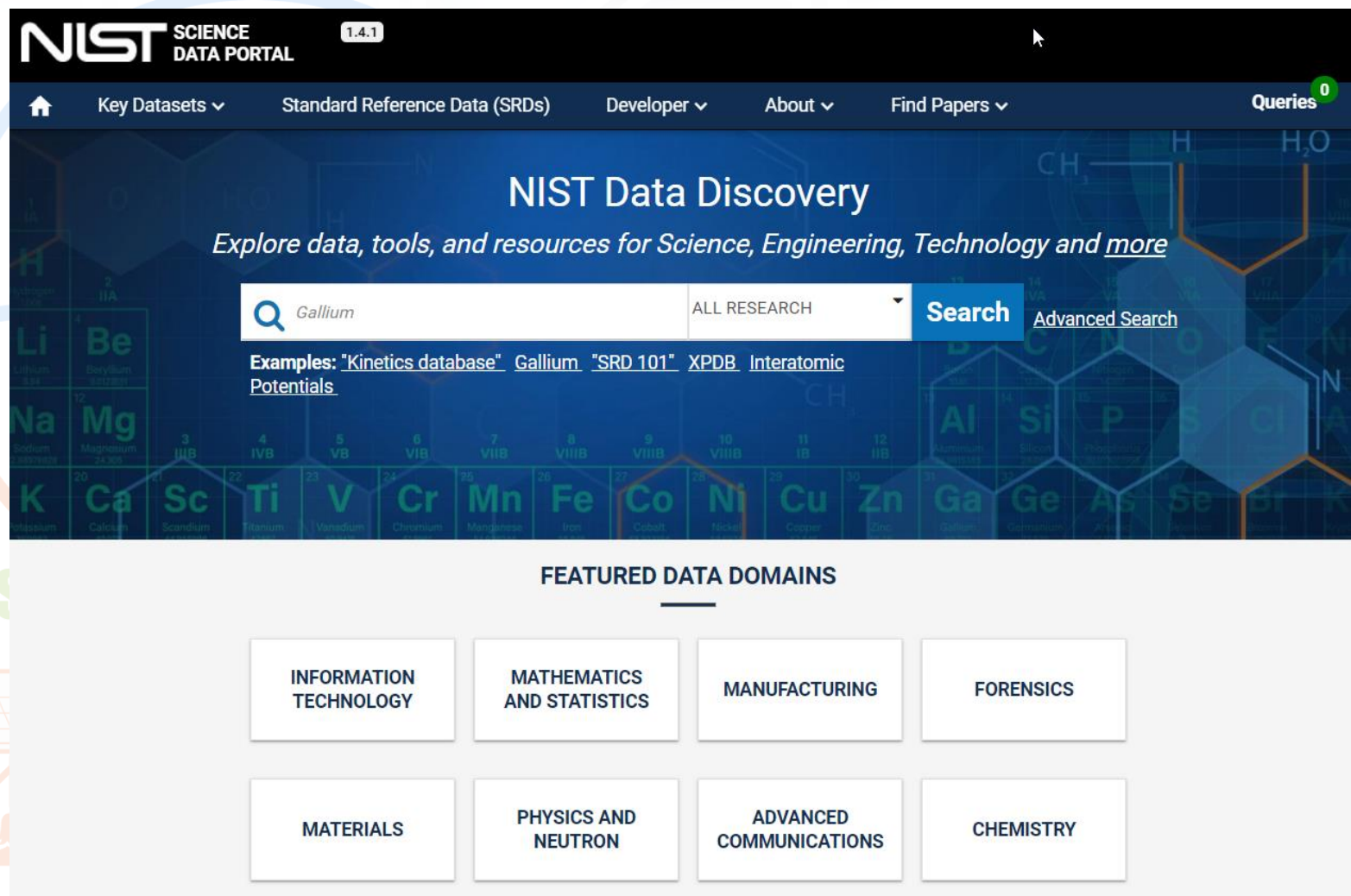


# Primary ODI Activities

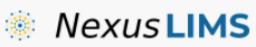
- Data management
  - Public Data Repository (PDR) and Science Data Portal (SDP), data.gov compliance
  - Laboratory Information Management Systems (LIMS)
  - Configurable Data Curation System (CDCS), Python-based metadata extractors (HyperSpy)
  - Data Management Plans (DMPs)
- Standard Reference Data (SRD)
- Informatics and analytics
- External engagements

# 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)



# Laboratory Information Management Systems



NexusLIMS

Q Browse and Search Records

Sharepoint Calendar

Tutorial

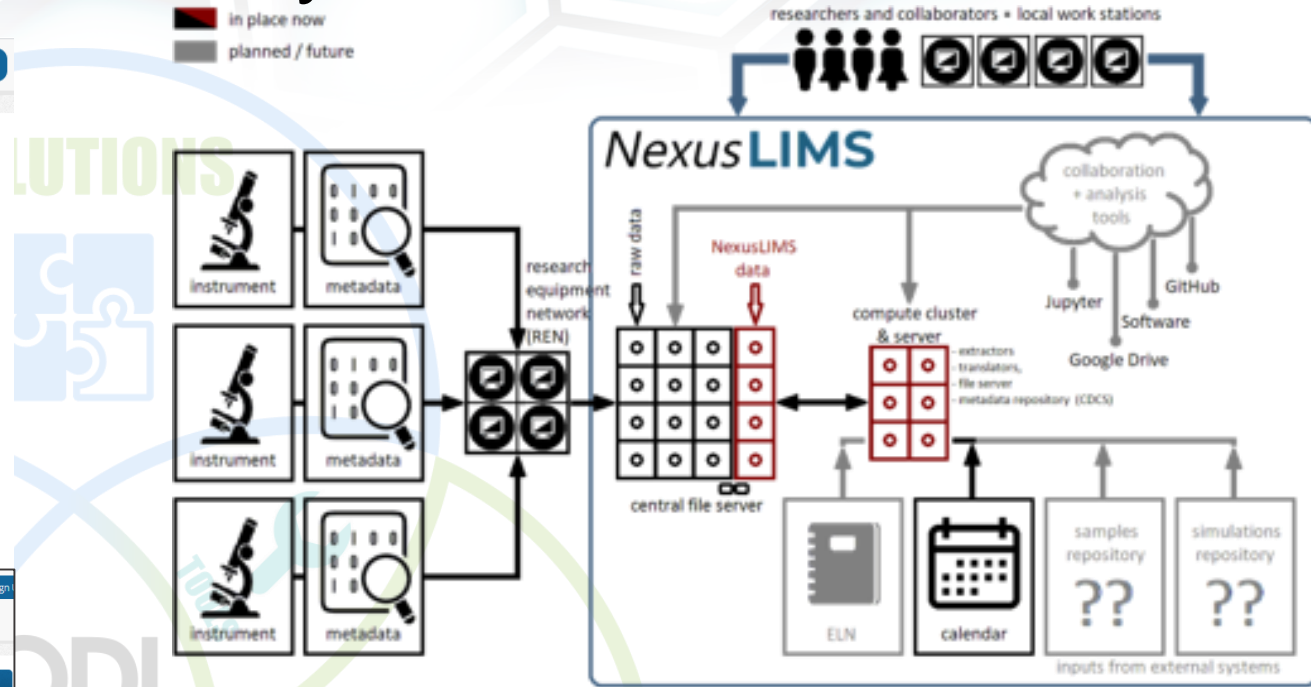
Help

Log In / Sign Up

## Welcome to NexusLIMS!

This laboratory information management system (LIMS) allows for the automated creation and curation of microscopy experimental records using the schema co-developed by ODI and the MML Electron Microscopy Nexus Facility. Experimental records are automatically harvested from multiple data sources to facilitate browsing and searching of data collected from the varied instruments in the Nexus Facility.

To learn more about how to use NexusLIMS, please take the [interactive tutorial](#), or visit the [documentation page](#). To get started, please click the link below to start browsing experimental records:



**Browse and Search Records**  
Click here to explore the NexusLIMS repository

NexusLIMS

Q Browse and Search Records

Sharepoint Calendar

Tutorial

Log In / Sign Up

Help

Search

stem

stem\_imaging

stem

stem\_eels

stem\_eds

Cur trenches with SHIMS

Raciti, David M. (Fed) - September 03, 2020

Motivation: Imagine of filled Cu patterns

Electrodeposition of magnetic thin films.

Rus, Eric D. (Fed) - September 01, 2020

Motivation: To determine the composition of electrochemically deposited platinum nickel thin films

HAADF STEM

Johnston-Peck, Aaron C. (Fed) - September 01, 2020

Motivation: HAADF STEM on absorption of RE on Al2O3

4D STEM

Johnston-Peck, Aaron C. (Fed) - August 31, 2020

Motivation: 4D STEM on QW sample

Analysis of Organic Films

FEI Titan STEM 38 data files in 10 activities 26 dm3 12 dm4

Herzing, Andrew A. (Fed) - August 25, 2020

Motivation: LAADF STEM morphological characterization STEM-EELS compositional measurement

Session Summary

Date: 2020-08-25

Start Time: 00:00:00

End Time: 23:59:00

Session ID: 218

Sample name: Thin Organic Films

Sample ID: bfb37f70-e71c-42d0-8523-004ea306490a

Dataset 1 of 38

Activity 1 of 10

NIST PUBLIC DATA REPOSITORY

Data Resource

Nexus-Experiment: an XML schema for describing data collected from electron microscopes

Raymond L. Plante, Joshua A. Tallon, June W. Lau, Gretchen Greene, Marcus Newrock

Contact: Raymond Plante

Identifier: doi:10.18434/022045

Version: 1.0.0 Last modified: 2020-02-26

Description

We share an XML Schema for describing data collected and the phased process of data creation. It was developed in collaboration at NIST between the NIST Microscopy & Measurement Lab's Office of Data and Informatics. When automatically gathers metadata from the reservation call an XML document describing what was done. The system scientist-oriented summary of the microscopy experiment reference to samples). This schema is expected to be a global microscopy research community's schema.

Citation

Copy the recommended text to cite this resource

Plante, Raymond L., Tallon, Joshua A., Lau, June W., Greene, Gretchen, Newrock, Marcus (2020). Nexus-Experiment: an XML schema for describing data collected from electron microscopes. National Institute of Standards and Technology. <https://doi.org/10.18434/022045> (Accessed 2020-09-09)

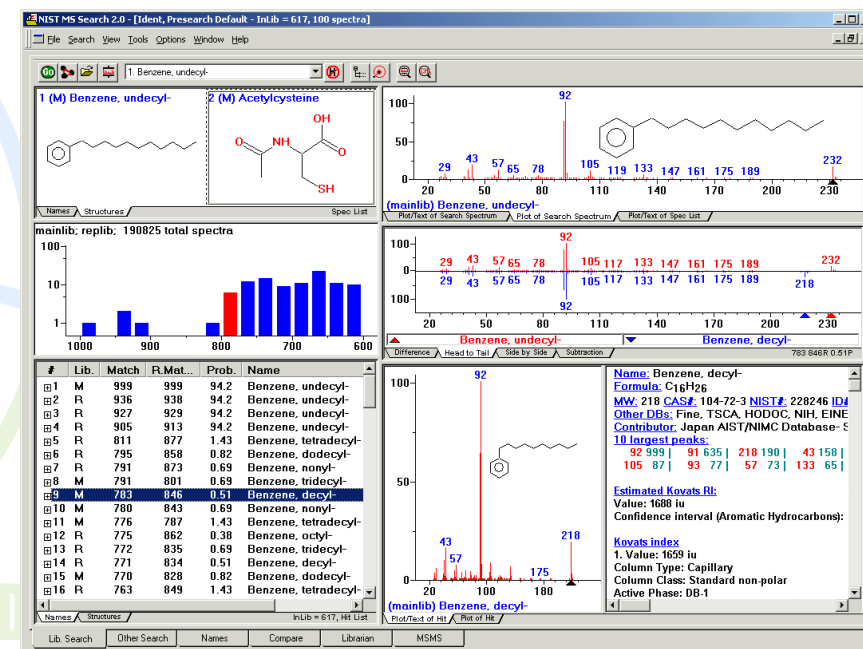
See also the NIST Citation Recommendations.

Research Topics: Metrology, Materials, Information Technology, Data and Informatics

Subject Keywords: XML Schema, data curation, metadata, microscopy, TEM, SEM, laboratory information management system (LIMS), laboratory notebook

# 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](http://webbook.nist.gov)
  - 300K views a month [XPS - NIST X-Ray Photoelectron Spectroscopy Database](#)



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)



# Informatics and Analytics Support

## Data Informatics Resources

[Python and R](#)

[AI and Machine Learning](#)

[Data Analytics and  
Uncertainty Quantification](#)

[Data Seminars and  
Training](#)

[Scientific computing](#)

## Data Informatics Resources

*A curated collection of data informatics references and learning materials relevant to NIST's mission in the materials, chemical, and biological sciences.*

### [Data Seminars and Training](#)

Information about Software/Data Carpentry, Python and R Slack channels, ODI seminars, and other events

### [Python and R](#)

Programming languages used widely in science and engineering

### [Artificial Intelligence and Machine Learning](#)

Instructional material related to applications of AI/ML in scientific research

### [Data analysis and uncertainty quantification](#)

References and resources, including some maintained by the NIST Statistical Engineering Division

### [Scientific computing links](#)

Includes high performance computing (Enki, Nisaba, etc.), scientific software, and data storage

# External Engagement

- Commerce Data Governance Board, Data Inventory Working Group
- OSTP/NSTC subcommittees (Subcommittee on Open Science, Subcommittee on Rigor and Integrity of Research)
- Research Data Alliance, CODATA (Digital Representation of Units of Measure task group), GO-FAIR (FAIR Digital Object Framework), World Data Service (Technical Advisory Board)
- Digital SI (BIPM/CIPM), Digital Calibration Certificates
- Commerce, Energy, NASA, Defense Information (CENDI) network
- National Academies Roundtable, Incentives for Open Data
- Association of American Universities (AAU) / Association of Public and Land-Grant Universities (APLU) / Association of Research Libraries (ARL) workshops on improving public access to research data
- Materials Research Data Council (MaRDaC) / Materials Research Data Network (MaRDaN)



# 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?

# Big Data Landscape 2016 (Version 3.0)

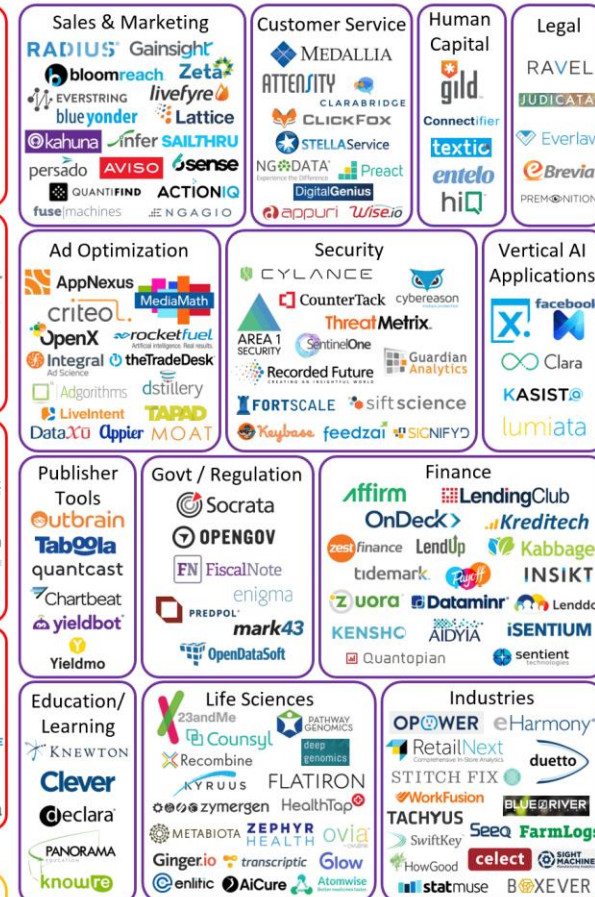
## Infrastructure



## Analytics



## Applications



## Cross-Infrastructure/Analytics



## Open Source



## Data Sources & APIs





# 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



## 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 Group



Laura Biven, DOE



Mercé Crosas, Harvard



Josh Greenberg, Sloan



Hilary Hanahoe, RDA



Heather Joseph, SPARC



Barend Mons, CODATA  
and GO-FAIR



Beth Plale, NSF



Anita de Waard, Elsevier



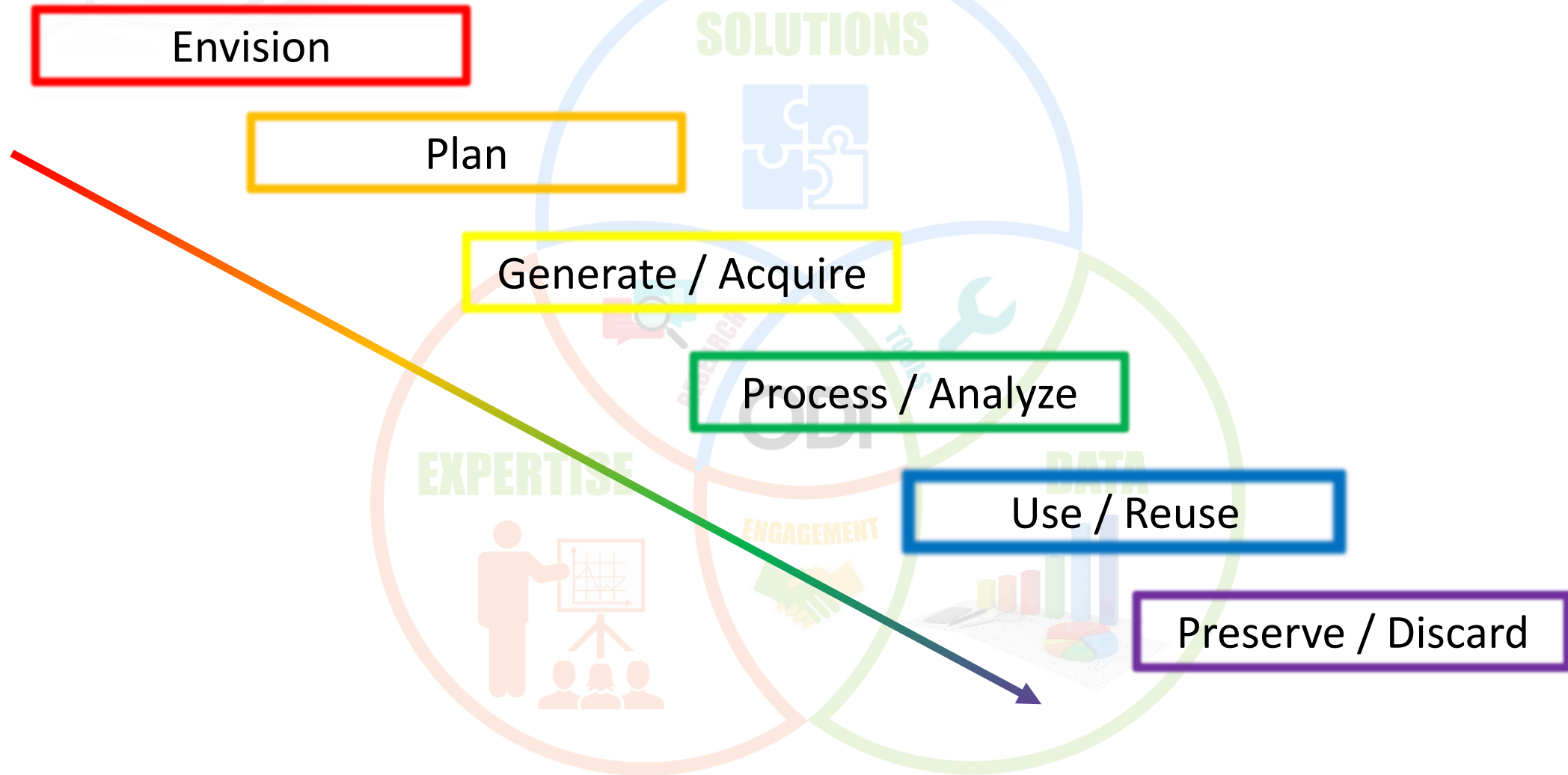
# Workshop Summary

- **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 Structure Based on “Functions”



# RDaF Structure

## Function

### 1) Envision

## Category

*Data governance*

*Community  
engagement*

*Data culture*

*Reward structure*

## Subcategory

Data vision, data policy  
Data management organization  
Data quality, privacy, ethics

Communication, interactions  
Cross-domain

FAIR principles  
Value of data  
Roles and responsibilities

Value of data professionals  
Incentives for sharing and re-use



# RDaF Structure

## Function

### 2) Plan

## Category

Costs

Funding

Data objects

Data management  
planning

## Subcategory

Cost-benefit analysis  
Costs by data lifecycle stage

Direct, overhead, mixed, other

Data (experimental, simulation)  
Software, instruments  
Publications, presentations

DMPs (intent, update)  
Formats, standards

# RDaF Structure

<u>Function</u>	<u>Category</u>	<u>Subcategory</u>
3) Generate / Acquire	Sources	In-house, experiment or simulation Collected from external sources
	Experiment	Instruments and their metadata Measurement protocols Data capture and recording
	Simulation	Commercial or custom software Metadata capture and recording
	External sources	Identification, provenance Metadata harvesting
	Data formats	Standards development and/or adoption

# RDaF Structure

## Function

### 4) Process / Analyze

## Category

*Provenance*

*Data architecture*

*Software*

*Publishing, curation*

## Subcategory

Origin, version, time-stamp

Data copied or derived from other data

Design, security, configuration management

Hosting and storage

On-premise or Cloud

Commercial or custom software

Versions

Stability, resilience, adaptability, maintenance

Workflows, ELNs, LIMs

Processes, tools, stewardship

Metadata



# 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

*Analysis tools*

AI/ML  
Performance

*Impact*

Usage tracking, citation

# 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)  
Migration between organizations

*Retention and disposition*

Decision processes  
End-of-life (dark archives, deaccession, gravestones)

# Status

- Briefed OSTP Subcommittee on Open Science and OSTP Director Kelvin Droegemeier (03/26/2020)



- Developed roadmap and structure, vetted with Steering Group
- Seeking ~\$500k to fund two pilots: materials science and research universities/libraries/scholarly publishers
  - NIST plus other agencies/laboratories, either \$\$ or in-kind support
  - Professional societies
  - Scholarly publishing community

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



# 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)

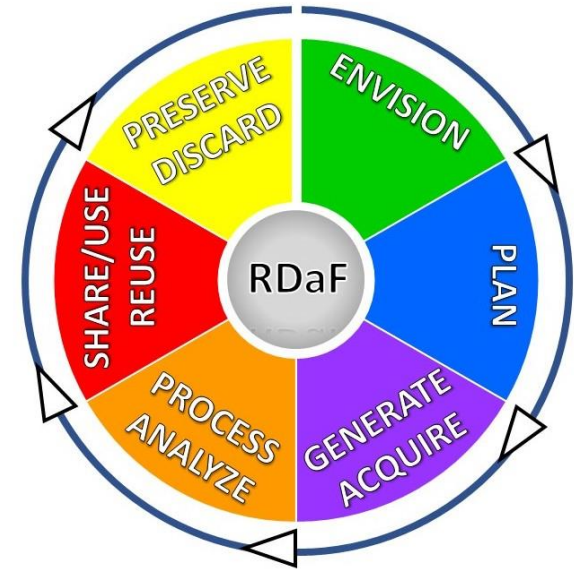
[https://www.nist.gov/system/files/documents/2019/09/09/nist\\_privacy\\_framework\\_preliminary\\_draft.pdf](https://www.nist.gov/system/files/documents/2019/09/09/nist_privacy_framework_preliminary_draft.pdf)

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

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

# 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
  - Start with pilot projects in order to validate approach and re-tune as necessary
  - Collaborate with other federal agencies, professional societies, scholarly publishing community, etc., to garner the necessary resources and take advantage of work in progress



# Contact

## SOLUTIONS

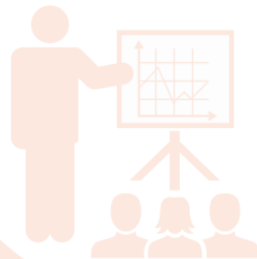
Robert Hanisch

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

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



## EXPERTISE



## DATA

