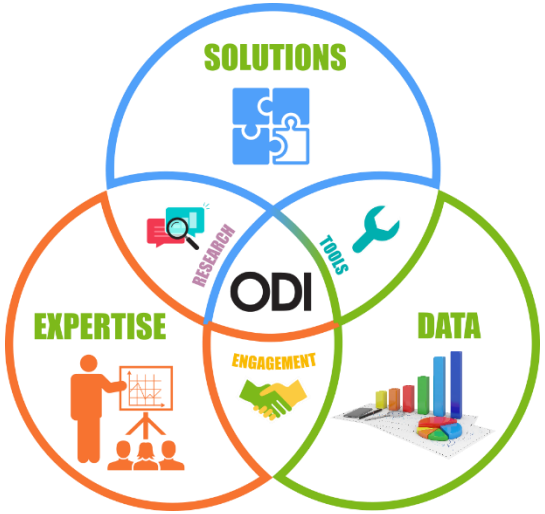
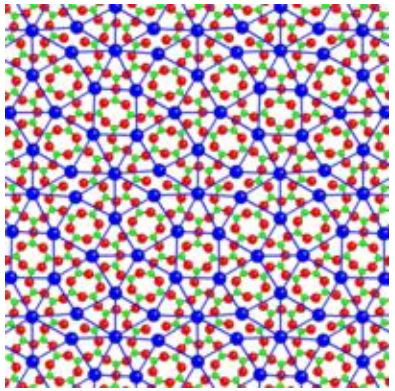
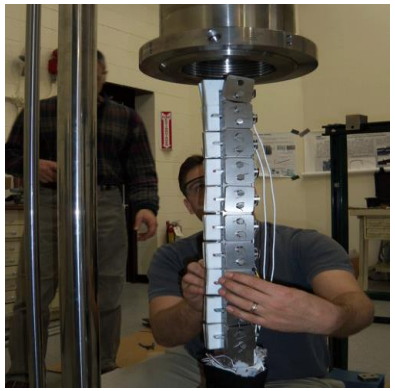
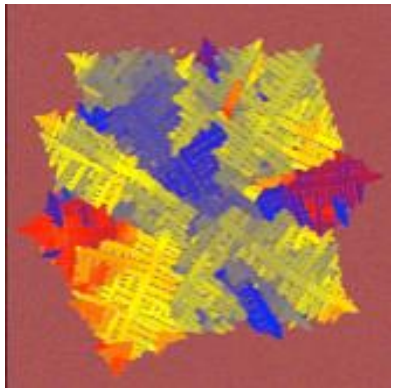








NIST Research Data Framework (RDaF)

Stakeholder Workshop on <<job/role>>

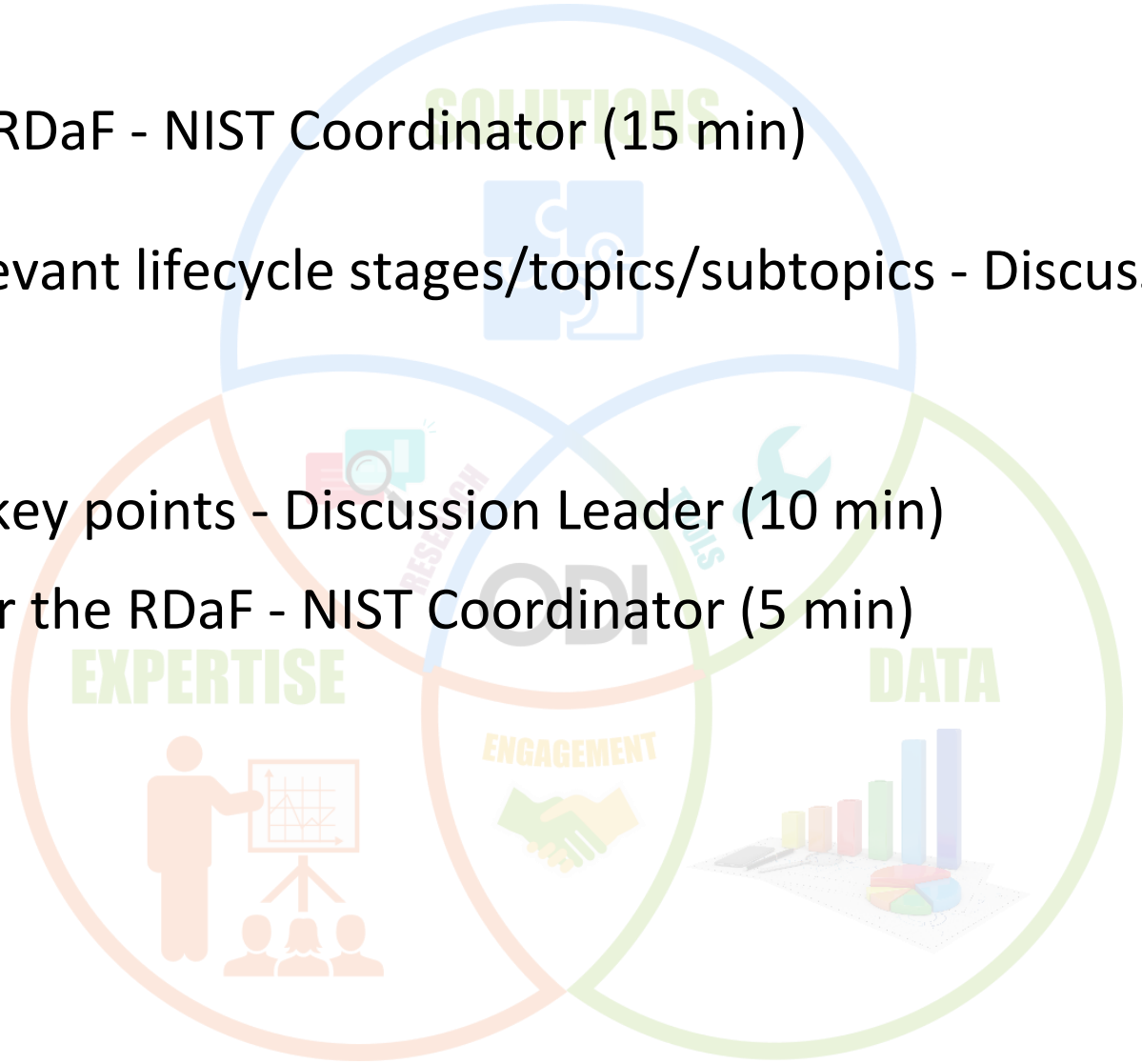
Discussion Leader:

NIST Coordinators:

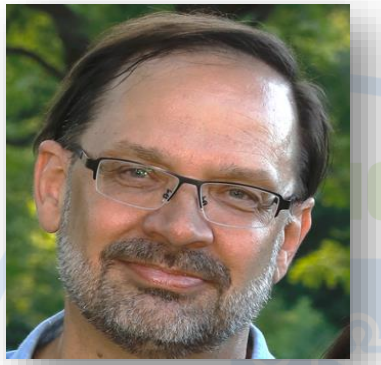
This workshop is being recorded

Stakeholder Workshops Agenda

- Overview of the RDaF - NIST Coordinator (15 min)
- Discussion of relevant lifecycle stages/topics/subtopics - Discussion Leader (90 min)
- Wrap-Up
  - Summary of key points - Discussion Leader (10 min)
  - Next steps for the RDaF - NIST Coordinator (5 min)



NIST Research Data Framework (RDaF) Team



Robert Hanisch
Director, ODI
RDaF Project Manager

Bonnie Carroll
Strategic Consultant
Steering Committee Chair

Debbie Kaiser RDaF
Senior Advisor, ODI
Deputy Project Manager

Andrea Medina-Smith
Research Librarian, NIST
Information Coordinator

Alda Yuan
Physical Scientist, ODI
Implementation Lead

Eva Campo
RDA US
Participant Interactions

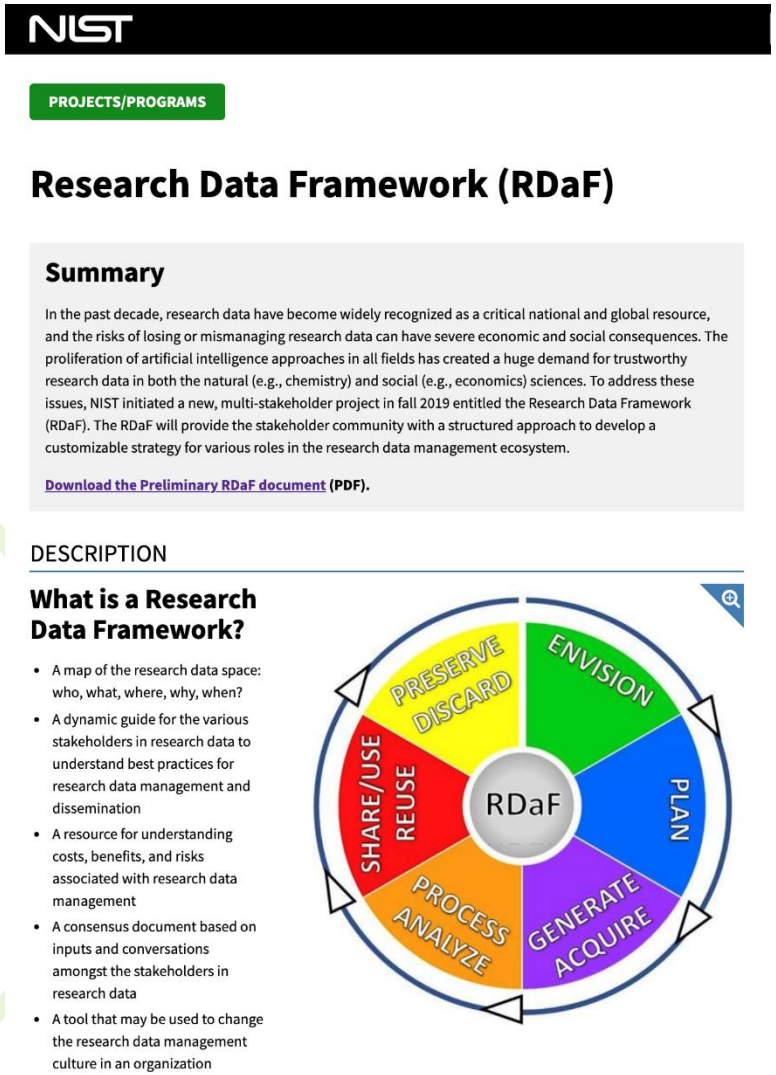
Amy Thompson
Executive Assistant
CampoStella Research & Consulting, LLC

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?



RDaF Publication

Framework Core, v. 1.0

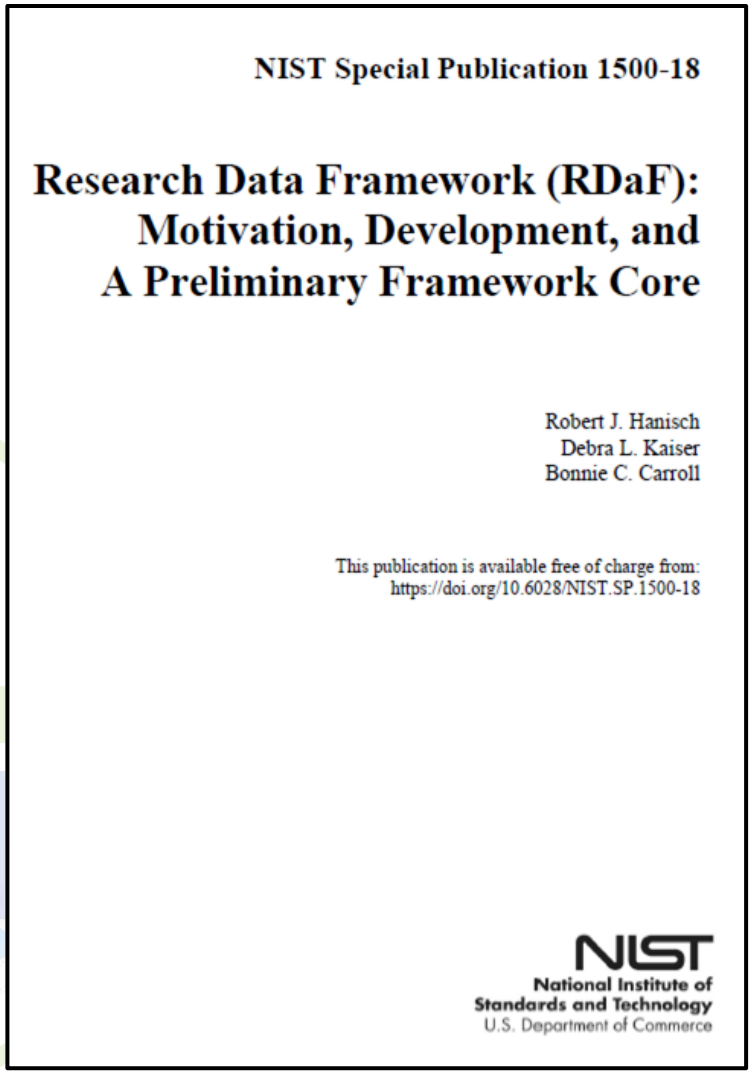
Six Research Data Lifecycle Stages

- **Topics**
- **Subtopics**
- Informative References

Research Data Lifecycle Stage	Topic	Subtopic
Envision		
Plan		

"A large spreadsheet"

Published February 2021



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

PLAN Topics and Subtopics ver. 1.1

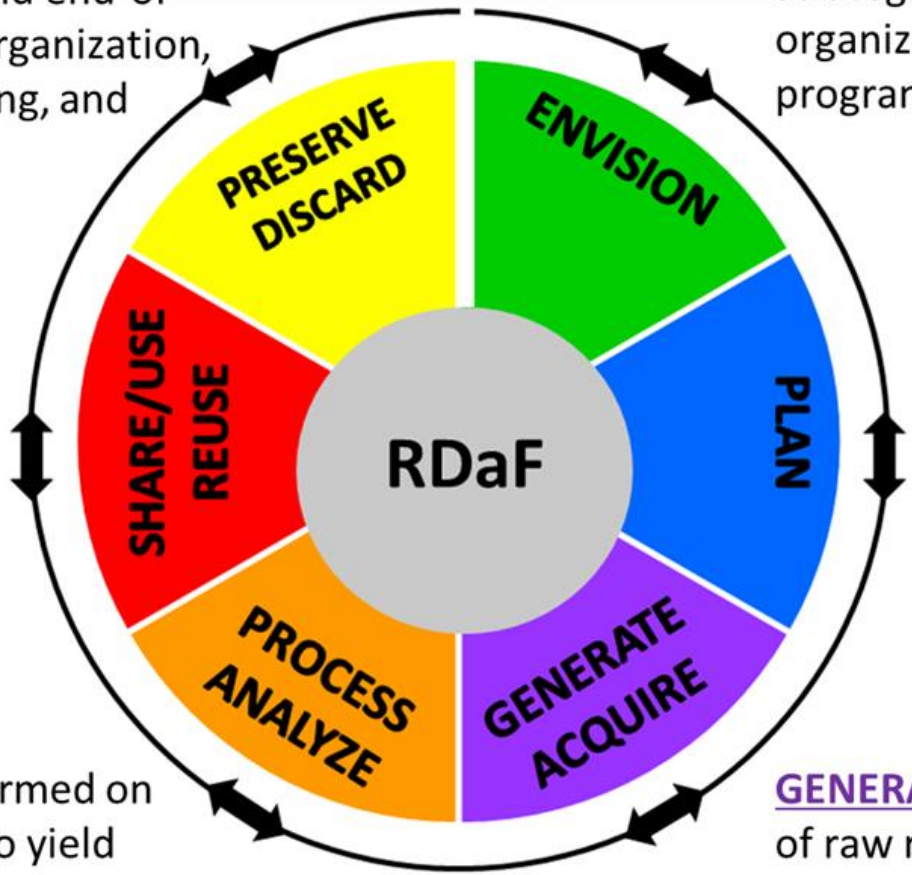
Topics	Subtopics
Chain of Control (Custody of Data)	• Roles and responsibility • Implementation authority • Communication and coordination within an organization • Documentation • Provenance
Financial Aspects of Planning	• Decision-making tools to assess costs of data • Cost-benefit analysis • Cost breakdown by lifecycle stage • Funding models for provisioning resources • Funding sources • Downstream lifecycle costs
Data Management Planning	• Written data management plans (DMPs) • Purpose/intent of research study • Specification of data objects, metadata, analysis tools, and workflows throughout the lifecycle • Lifecycle considerations: living documents with regular review or static proposals? • Machine-readable DMPs • Linkage of DMPs to other administrative records • Data organization (e.g., database, repository) to facilitate future access
Data Objects	• Data: quantitative and qualitative • Software • Models • Observational Data • Survey Data • Documentation (text) • Specimens (physical samples) • Presentations
FAIR	• Identification of methods/guidelines vis-a-vis FAIR principles • Organizational support for making data FAIR

Topics	Subtopics
Data Architecture	• Design • Data processing operations • Data flows • Data models • Laboratory information management systems • Hosting and storage, cloud storage • Configuration management • Interoperability with other architectures • Security • Existing standards
Data/Metadata Considerations	• Criteria for selection of data/metadata • Nature of data/metadata required • Intended extent of FAIRness • Methods to capture and store data/metadata • Metadata schema
Research Data Standards	• Needs for standards • Criteria: general vs. domain-specific standards • Sources of standards/guidelines for data/metadata • Data quality standards • De facto standards/conventions (community-based)
Hardware and Software Infrastructure	• Anticipation of needs • Tools to support data-related processes • Models that connect infrastructure to data processes and workflow • Interoperability • Persistent instrument identifiers • Sustainability of data vis-à-vis obsolete infrastructure • Security and privacy considerations

RDaF Lifecycle Stages

PRESERVE/DISCARD: The end-of-use and end-of-life provisions for research data in an organization, including records management, archiving, and safe disposal.

ENVISION: Review of the overall strategies and drivers of an organization's research data program.



SHARE/USE/REUSE: How research data are disseminated, used, and reused within and outside an organization.

PLAN: The tactical management positioning in an organization for effective research data management throughout the research data lifecycle.

PROCESS/ANALYZE: The actions performed on generated or acquired research data to yield processed research data, and the research data stewardship functions performed by an organization.

GENERATE/ACQUIRE: The generation of raw research data and/or the acquisition of research data by an organization

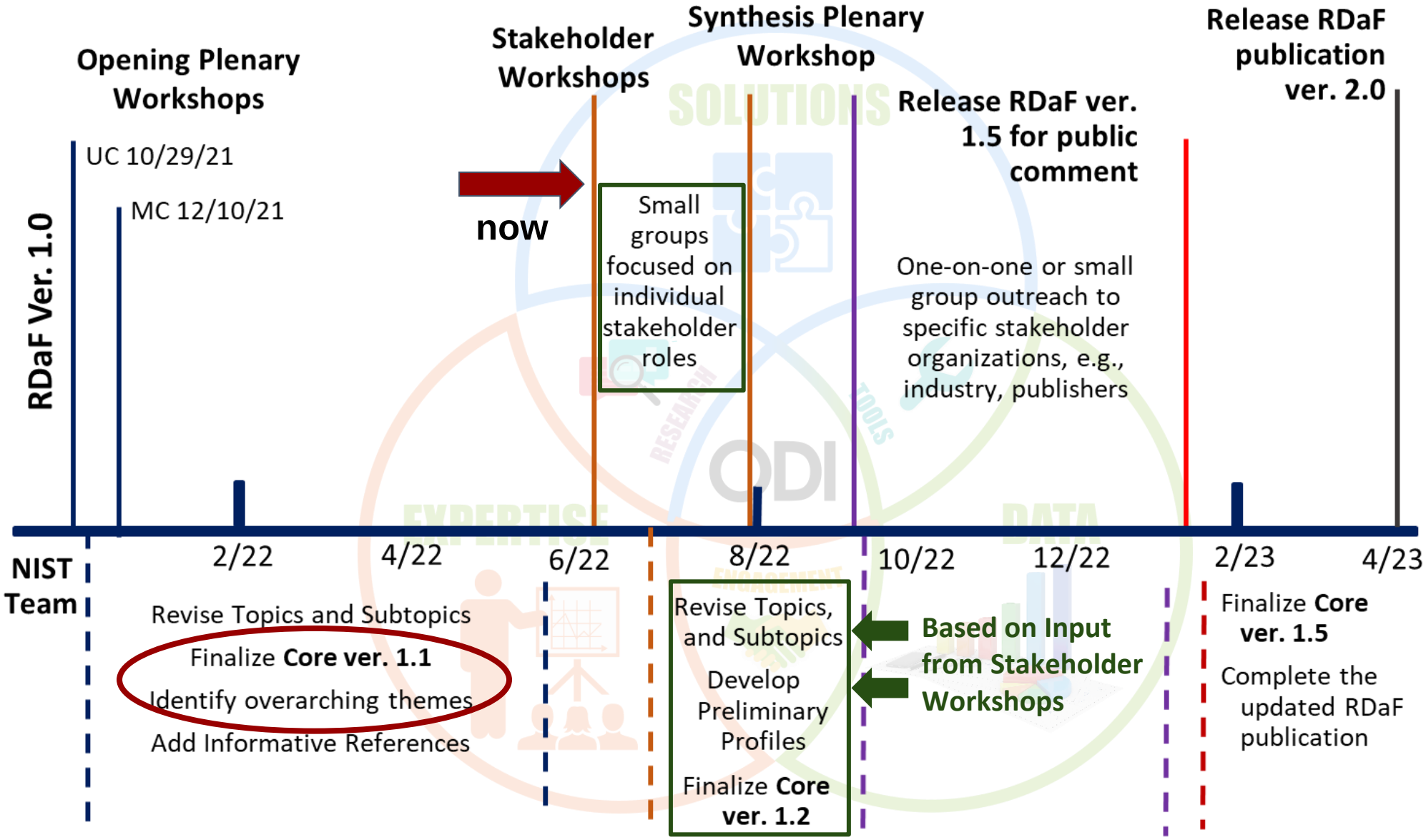
Roles for Stakeholder Workshops

- Academic Mid-Execs/Heads of Research
- AI Experts
- Budget/Cost Experts
- Curators
- Data/IT Leaders
- Data/Research Governance Leaders
- Funders
- Institute/Center/Program Directors
- Open Data Experts
- Professional Society/Trade Organization Leaders
- Professors
- Providers of Data Tools/Services/ Infrastructure
- Publishers
- Researchers
- Senior Executives

- These roles help RDaF users select the parts of the Framework Core that are most important for them
- The series of workshops will help the NIST team understand how these roles interact with one another in the context of the broad data ecosystem

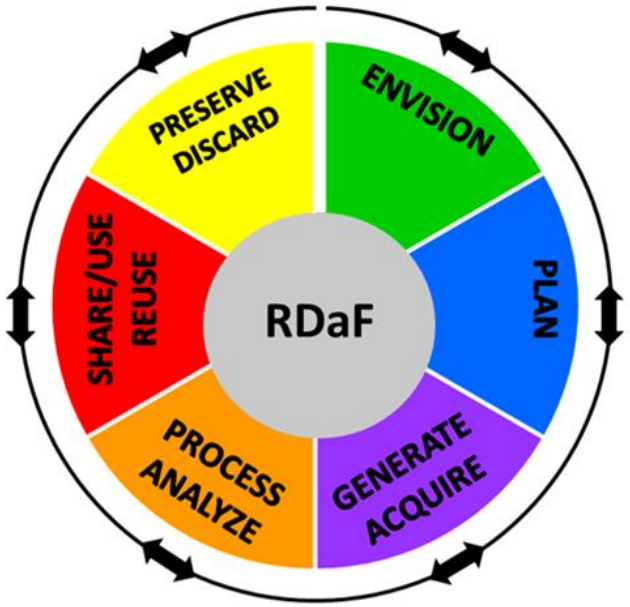
Return to Stages

RDaF Development Timeline



RDaF Profiles

- Facilitate use of the RDaF by individuals in various job roles within an organization
- Serve as a checklist of lifecycle stages/topics/subtopics relevant to a job role ***with respect to research data and research data management***
- NIST team will generate "Generic" Profiles as starting points using input from the workshops
- These "Generic" Profiles can be refined to fit an individual's role(s)



Example of a partial profile for three different roles

			Roles		
Research Data Lifecycle Stage	Topic	Subtopic	Senior Executive	Researcher	Curator
ENVISION Review of the overall strategies and drivers of an organization's research data program.	Data Governance: Strategic/ Qualitative	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		

Workshop Objectives

- Help develop profiles for various roles to support the use of the Framework.
- Foster an open discussion to identify the most relevant research data lifecycle stages/topics/subtopics for <<job/role>>
- Work from the research lifecycle stages to topics to subtopics

Are there any missing topics or subtopics? Are any misplaced?

- In other words,
  - What parts of the framework influence your job?
  - What parts of the framework do you influence?

Discussion on <<*Job Role*>>

Discussion Leader:

NIST Coordinators:

Contact Name	Affiliation	Title

Discussion Agenda

- Introductions
  - Name, job title, organization
  - Which of the six Research Data Lifecycle Stages are relevant for your job role? Which do you have influence over?
- Based on the group's priority Stages, select those to discuss in more detail
  - *Priority Stage 1*
    - Topics
    - Subtopics for the group's priority topics
  - *Priority Stage 2*
    - Topics
    - Subtopics for the group's priority topics
  - *Priority Stage 3* (time permitting)…

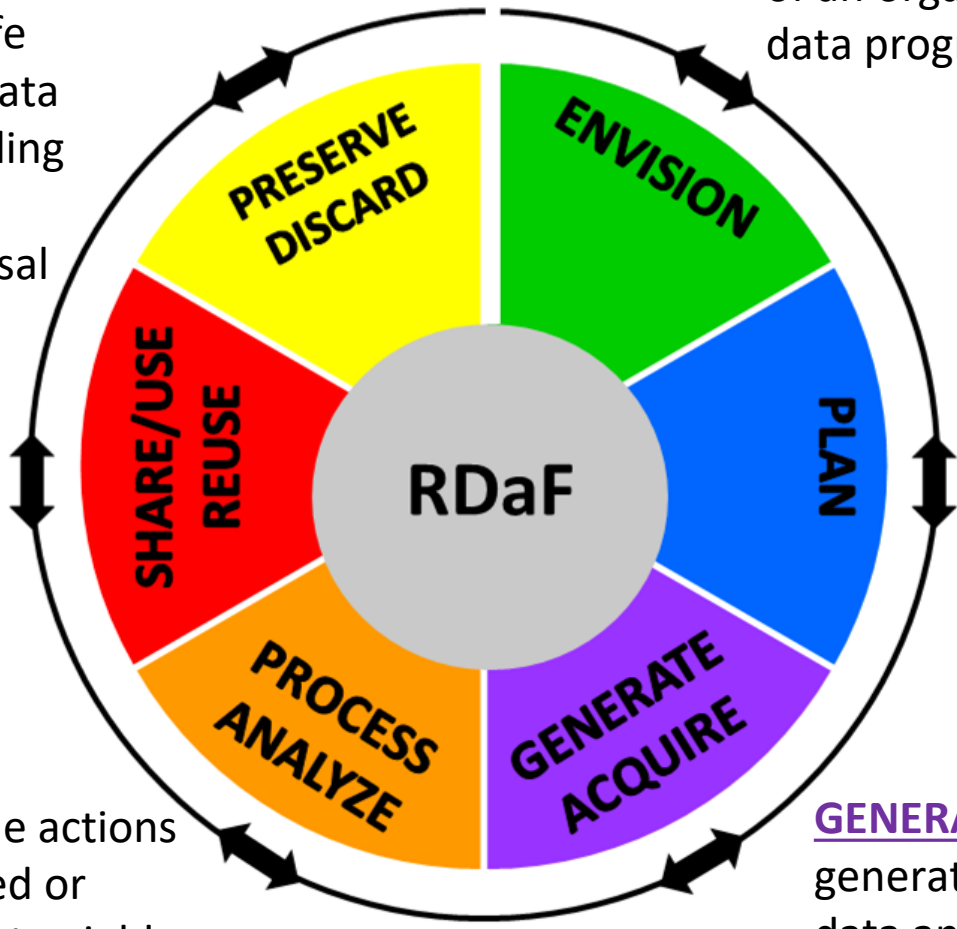
Research Data Lifecycle Stages

For your role as a <<job/role>>, *with respect to research data and the management of research data*:

- **Which lifecycle stages are relevant?**
- **Which stages influence your role?**
- **Which stages do you influence?**

PRESERVE/DISCARD: The end-of-use and end-of-life provisions for research data in an organization, including records management, archiving, and safe disposal

ENVISION: Review of the overall strategies and drivers of an organization's research data program



SHARE/USE/REUSE: How research data are disseminated, used, and reused within and outside an organization

PLAN: The tactical management positioning in an organization for effective research data management throughout the research data lifecycle

PROCESS/ANALYZE: The actions performed on generated or acquired research data to yield processed research data, and the research data stewardship functions performed by an organization

GENERATE/ACQUIRE: The generation of raw research data and/or the acquisition of research data by an organization

Other roles

Next Steps

ENVISION Lifecycle Stage

Data Governance—Strategic/Qualitative

- Identification of Goals and Roles
- Data vision and/or data policy
- Data management organization
- Organizational values
- Data management value proposition
- Data needs assessment
- Data quality
- Value of data (quantitative or qualitative)
- Organization intent regarding FAIR data
- Data end-use support
- Data stewardship
- Coordination and information sharing across an organization

Education and Workforce Development

- Workforce skills inventory
- HR's role in data workforce development
- Data management training
- Workforce preparedness in new and advancing technologies, e.g., HPC, AI, ML, and computation services
- Promotional paths, continual training, and career development

Data Governance—Legal and Regulatory Compliance

- Data services and the beneficiaries
- Data privacy
- Data ethics
- Safety and security assurance
- Data inventory
- Risk assessment
- Risk mitigation and management
- Data sharing/licensing
- Social license for the use of data
- Jurisdiction for data storage

Community Engagement

- Stakeholder community(ies): internal and external
- Communication with stakeholder community(ies)
- Interactions with other organizations
- Cross-community engagement (across domains and sectors)
- Inclusivity in interactions
- Partners/Partnerships

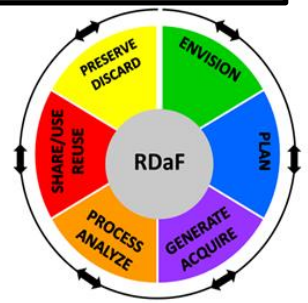
Data Culture and Reward Structure

- Roles and responsibilities
- Value of data to organization and leadership
- Recognition of data management
- Value of data workers
- Promotion and tenure
- FAIR data principles
- Maintaining FAIR data
- Incentives and institutional credit for data sharing and reuse
- Disincentives for data sharing
- CARE*

Resources— Allocation and Sustainability

- Sources of funding
- Long-term funding

*Collective Benefit, Authority to Control, Responsibility, and Ethics



ENVISION: Topics

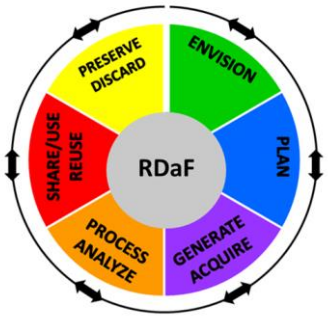
- Data Governance—Strategic/Qualitative
- Data Governance—Legal and Regulatory Compliance
- Community Engagement
- Data Culture and Reward Structure
- Resources—Allocation and Sustainability
- Education and Workforce Development

Which topics are relevant to your job role or do you have influence over?

Are some topics more important than others?

Are any topics missing?

Return to ENVISION Main Page



ENVISION: Data Governance—Strategic/Qualitative Subtopics

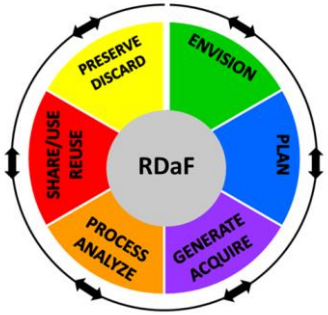
- Identification of Goals and Roles
- Data vision and/or data policy
- Data management organization
- Organizational values
- Data management value proposition
- Data needs assessment
- Data quality
- Value of data (quantitative or qualitative)
- Organization intent regarding FAIR data
- Data end-use support
- Data stewardship
- Coordination and information sharing across an organization

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to ENVISION Main Page



ENVISION: Data Governance—Legal and Regulatory Compliance Subtopics

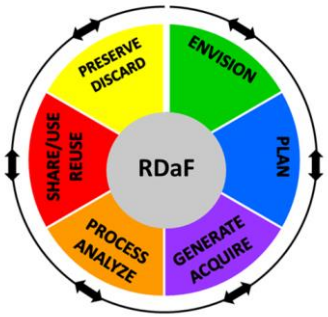
- **Data privacy**
- **Data ethics**
- **Safety and security assurance**
- **Data inventory**
- **Risk assessment**
- **Risk mitigation and management**
- **Data sharing/licensing**
- **Social license for the use of data**
- **Jurisdiction for data storage**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to ENVISION Main Page



ENVISION: Community Engagement Subtopics

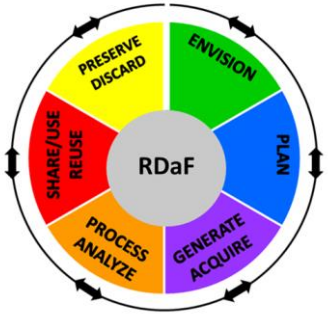
- **Stakeholder community(ies): internal and external**
- **Communication with stakeholder community(ies)**
- **Interactions with other organizations**
- **Cross-community engagement (across domains and sectors)**
- **Inclusivity in interactions**
- **Partners/Partnerships**
- **Data services and the beneficiaries**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to ENVISION Main Page



ENVISION: Data Culture and Reward Structure Subtopics

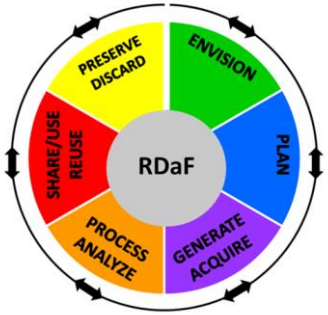
- **Roles and responsibilities**
- **Value of data to organization and leadership**
- **Recognition of data management**
- **Value of data workers**
- **Promotion and tenure**
- **FAIR data principles**
- **Maintaining FAIR data**
- **Incentives and institutional credit for data sharing and reuse**
- **Disincentives for data sharing**
- **CARE**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to ENVISION Main Page



ENVISION: Resources—Allocation and Sustainability Subtopics

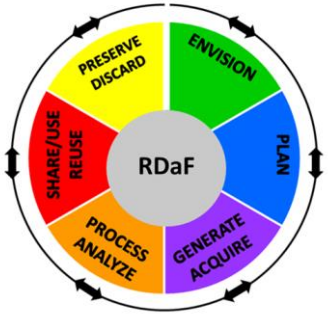
- **Sources of funding**
- **Long-term funding**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to ENVISION Main Page



ENVISION: Education and Workforce Development Subtopics

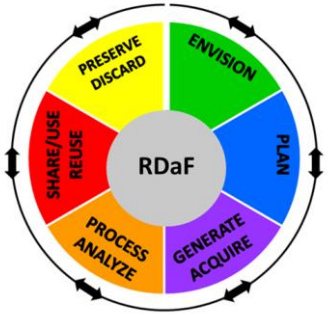
- **Workforce skills inventory**
- **Human Resource's role in data workforce development**
- **Data management training**
- **Workforce preparedness in new and advancing technologies, e.g., high-performance computing, AI, machine learning, and computation services**
- **Promotional paths, continual training, and career development**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to ENVISION Main Page



PLAN Lifecycle Stage (page 1)

Chain of Control (Custody of Data)

- Roles and responsibility
- Implementation authority
- Communication and coordination within an organization
- Documentation
- Provenance

Financial Aspects of Planning

- Decision-making tools to assess costs of data
- Cost-benefit analysis
- Cost breakdown by lifecycle stage
- Funding models for provisioning resources
- Funding sources
- Downstream lifecycle costs

Data Objects

- Data: quantitative and qualitative
- Software
- Models
- Observational Data
- Survey Data
- Documentation (text)
- Specimens (physical samples)
- Presentations

Data/Metadata Considerations

- Criteria for selection of data/metadata
- Nature of data/metadata required
- Intended extent of FAIRness
- Methods to capture and store data/metadata
- Metadata schema

Data Architecture

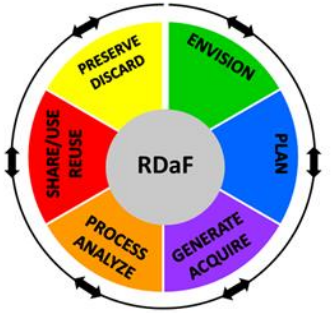
- Design
- Data processing operations
- Data flows
- Data models
- LIMS
- Hosting and storage, cloud storage
- Configuration management
- Interoperability with other architectures
- Security
- Existing standards

Assessment

- Goals/definition of success
- Metrics or metrics structure, tracking use and impact measures

Data Management Planning

- Written data management plans (DMPs)
- Purpose/intent of research study
- Specification of data objects, metadata, analysis tools, and workflows throughout the lifecycle
- Lifecycle considerations: living documents with regular review or static proposals?
- Machine-readable DMPs
- Linkage of DMPs to other administrative records
- Data organization (e.g., database, repository) to facilitate future access



PLAN Lifecycle Stage (page 2)

Hardware and Software Infrastructure

- Anticipation of needs
- Tools to support data-related processes
- Models that connect infrastructure to data processes and workflow
- Interoperability
- Persistent instrument identifiers
- Sustainability of data vis-à-vis obsolete infrastructure
- Security and privacy considerations

Research Data Standards

- Needs for standards
- Criteria: general vs. domain-specific standards
- Sources of standards/guidelines for data/metadata
- Data quality standards
- De facto standards, conventions (community-based)

FAIR

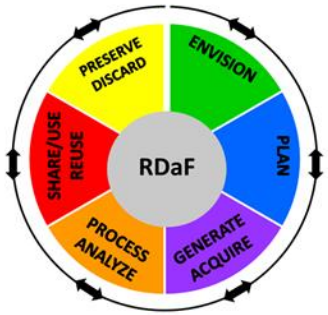
- Identification of methods/guidelines vis-a-vis FAIR principles
- Organizational support for making data FAIR

Communication and Outreach

- Methods to share data/metadata
- Promotion of data to communities of interest
- Requests for additional data from community

Access Control Associated with Data Sensitivity

- Ease of maintenance and implementation of records
- Limited disclosure (e.g., IP)
- Licensing data for reuse



PLAN: Topics

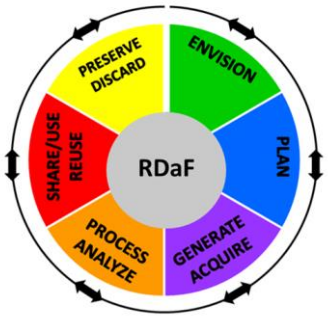
- Chain of Control (Custody of Data)
- Financial Aspects of Planning
- Data Management Planning
- Data Objects
- Data Architecture
- Data/Metadata Considerations
- FAIR
- Research Data Standards
- Hardware and Software Infrastructure
- Assessment
- Communication and Outreach
- Access Control Associated with Data Sensitivity

Which topics are relevant to your job role or do you have influence over?

Are some topics more important than others?

Are any topics missing?

Return to PLAN Main Page



PLAN: Chain of Control (Custody of Data) Subtopics

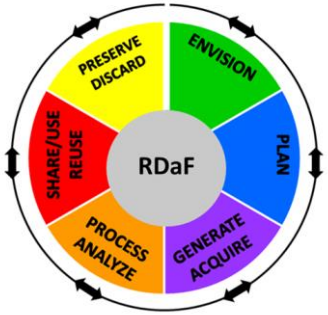
- **Roles and responsibility**
- **Implementation authority**
- **Communication and coordination within an organization**
- **Documentation**
- **Provenance**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Financial Aspects of Planning Subtopics

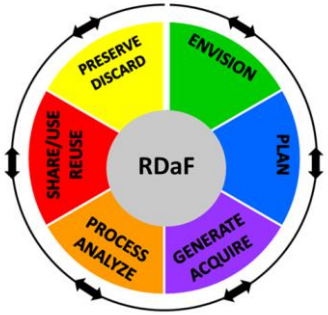
- **Decision-making tools to assess costs of data**
- **Cost-benefit analysis**
- **Cost breakdown by lifecycle stage**
- **Funding models for provisioning resources**
- **Funding sources**
- **Downstream lifecycle costs**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Data Management Planning Subtopics

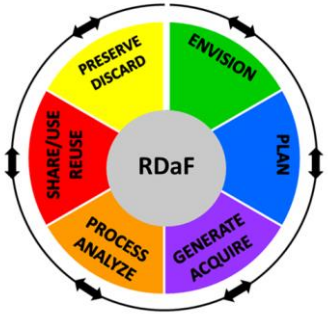
- **Written data management plans (DMPs)**
- **Purpose/intent of research study**
- **Specification of data objects, metadata, analysis tools, and workflows throughout the lifecycle**
- **Lifecycle considerations: living documents with regular review or static proposals?**
- **Machine-readable DMPs**
- **Linkage of DMPs to other administrative records**
- **Data organization (e.g., database, repository) to facilitate future access**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Data Objects Subtopics

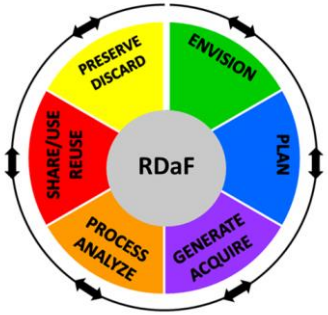
- **Data: quantitative and qualitative**
- **Software**
- **Models**
- **Observational Data**
- **Survey Data**
- **Documentation (text)**
- **Specimens (physical samples)**
- **Presentations**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Data Architecture Subtopics

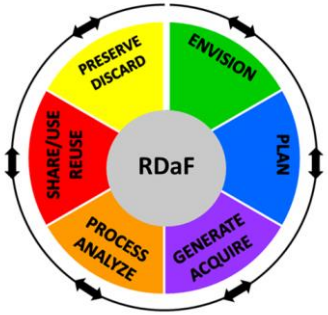
- **Design**
- **Data processing operations**
- **Data flows**
- **Data models**
- **Laboratory Information Management Systems**
- **Hosting and storage, cloud storage**
- **Configuration management**
- **Interoperability with other architectures**
- **Security**
- **Existing standards**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Data/Metadata Considerations Subtopics

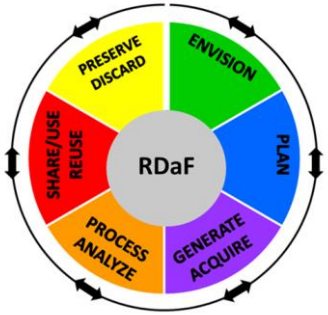
- **Criteria for selection of data/metadata**
- **Nature of data/metadata required**
- **Intended extent of FAIRness**
- **Methods to capture and store data/metadata**
- **Metadata schema**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: FAIR Subtopics

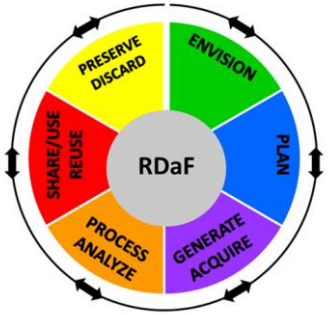
- **Identification of methods/guidelines vis-à-vis FAIR principles**
- **Organizational support for making data FAIR**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Research Data Standards Subtopics

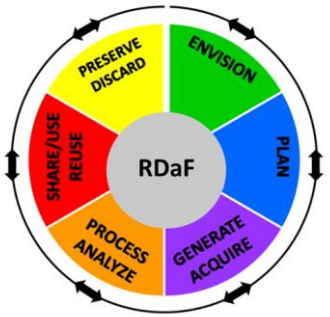
- **Needs for standards**
- **Criteria: general vs. domain-specific standards**
- **Sources of standards/guidelines for data/metadata**
- **Data quality standards**
- **De facto standards/conventions (community-based)**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Hardware and Software Infrastructure Subtopics

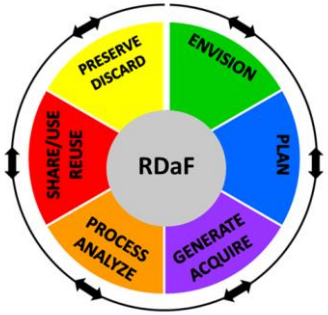
- **Anticipation of needs**
- **Tools to support data-related processes**
- **Models that connect infrastructure to data processes and workflow**
- **Interoperability**
- **Persistent instrument identifiers**
- **Sustainability of data vis-à-vis obsolete infrastructure**
- **Security and privacy considerations**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Assessment Subtopics

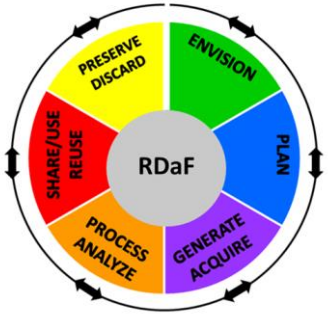
- **Goals/definition of success**
- **Metrics or metrics structure, tracking use and impact measures**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Communication and Outreach Subtopics

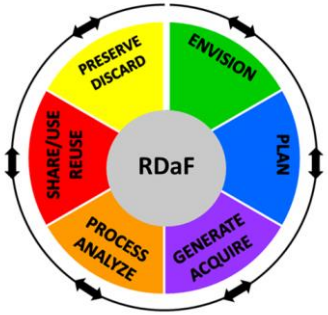
- **Methods to share data/metadata**
- **Promotion of data to communities of interest**
- **Requests for additional data from community**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



PLAN: Access Control Associated with Data Sensitivity Subtopics

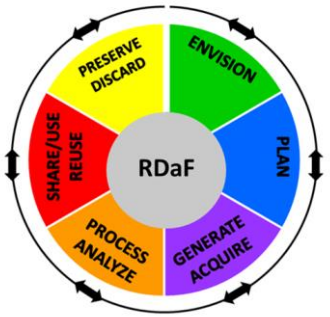
- **Ease of maintenance and implementation of records**
- **Limited disclosure (e.g., IP)**
- **Licensing data for reuse**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PLAN Main Page



GENERATE/ACQUIRE Lifecycle Stage

Types of Raw Data
- Measurements, including images, audio recordings, photos/videos
- Text files
- Computations, simulations
- Source code
- Observational data
- Survey data
- Other

Sources of Data
- In-house generation by researchers
- Remote generation by researchers
- In-field generation by researchers
- User facility generation for researcher

Acquired Data/Metadata
- Collaborators' data
- Data in repositories
- Aggregated datasets from multiple sources
- Extraction of data from the literature
- Provenance
- Restrictions
- Fees
- Usage agreements

Non-Computational Data/Metadata
- Source of object/subject
- Characteristics of object/subject
- Conditions of research study
- Instruments and other tools
- Methods and protocols
- Data/metadata capture methods
- Provenance and capture methods
- Reproducibility

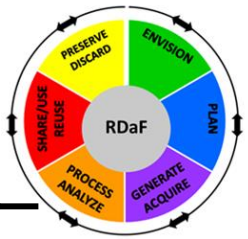
Computational Data/Metadata
- Calculations, models, simulation
- Input data/metadata
- Output data/metadata
- Hardware
- Data/metadata capture methods
- Provenance and capture methods
- Verification/validation of output data

Critically Evaluated (CE) Data
- Infrastructure to determine data quality
- Single researcher dataset
- Aggregation of data evaluated by experts
- Uncertainty quantification
- Modeling based on CE data

FAIR Principles
- Data born FAIR
- Data made FAIR
- Not FAIR data, e.g., legacy
- Guidelines/methodologies to assure or attain each aspect F, A, I, R
- Provenance or verification process for FAIR-ness

Community-Based Standards
- Data format and file structure
- Metadata format and file structure
- Vocabulary and ontology
- Interoperability
- Standards development organizations vs. grass roots efforts
- General vs. domain-specific

Acquisition Software
- Electronic laboratory notebooks
- Laboratory information management systems
- Instrument control
- Open source vs. proprietary



GENERATE/ACQUIRE: Topics

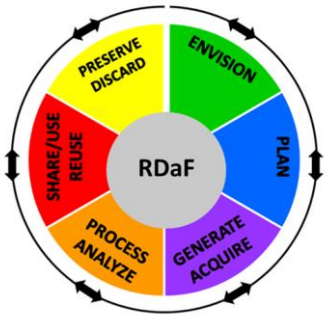
- Types of Raw Data
- Sources of Data
- Non-Computational Data/Metadata
- Computational Data/Metadata
- Critically Evaluated (CE) Data
- Acquired Data/Metadata
- FAIR Principles
- Community-Based Standards
- Acquisition Software

Which topics are relevant to your job role or do you have influence over?

Are some topics more important than others?

Are any topics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Types of Raw Data Subtopics

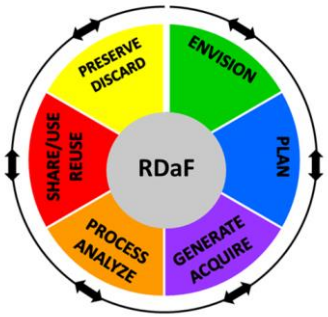
- **Measurements, including images, audio recordings, photos/videos**
- **Text files**
- **Computations, simulations**
- **Source code**
- **Observational data**
- **Survey data**
- **Other**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Sources of Data Subtopics

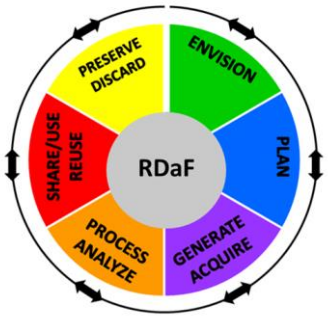
- **In-house generation by researchers**
- **Remote generation by researchers**
- **In-field generation by researchers**
- **User facility generation for researcher**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Non-Computational Data/Metadata Subtopics

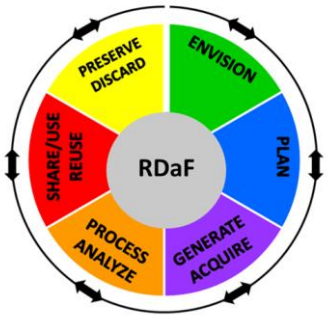
- **Source of object/subject**
- **Characteristics of object/subject**
- **Conditions of research study**
- **Instruments and other tools**
- **Methods and protocols**
- **Data/metadata capture methods**
- **Provenance and capture methods**
- **Reproducibility**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Computational Data/Metadata Subtopics

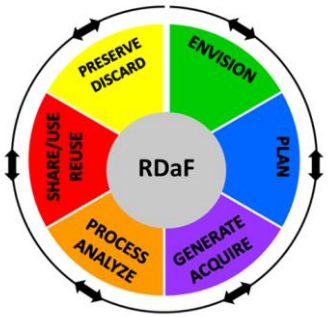
- **Calculations, models, simulations**
- **Input data/metadata**
- **Output data/metadata**
- **Hardware**
- **Data/metadata capture methods**
- **Provenance and capture methods**
- **Verification/validation of output data**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Critically Evaluated (CE) Data Subtopics

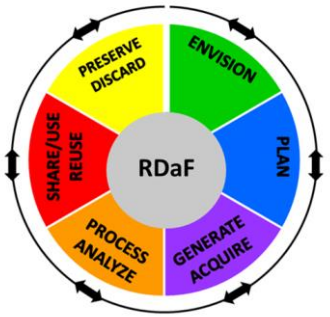
- **Infrastructure to determine data quality**
- **Single researcher dataset**
- **Aggregation of data evaluated by experts**
- **Uncertainty quantification**
- **Modeling based on CE data**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Acquired Data/Metadata Subtopics

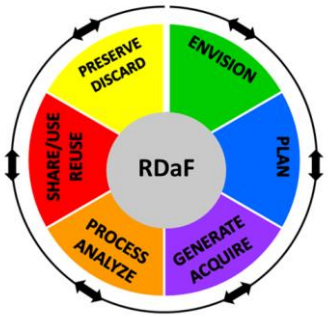
- **Collaborators' data**
- **Data in repositories**
- **Aggregated datasets from multiple sources**
- **Extraction of data from the literature**
- **Provenance**
- **Restrictions**
- **Fees**
- **Usage agreements**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: FAIR Principles Subtopics

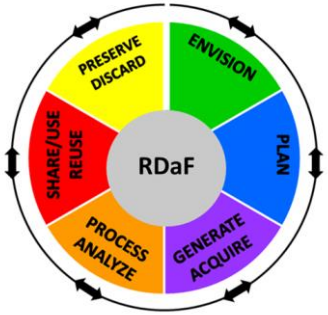
- **Data born FAIR**
- **Data made FAIR**
- **Not FAIR data, e.g., legacy**
- **Guidelines/methodologies to assure or attain each aspect F, A, I, and R**
- **Provenance or verification process for FAIR-ness**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Community-Based Standards Subtopics

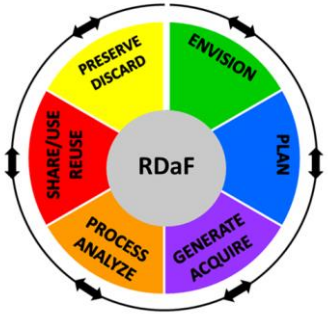
- **Data format and file structure**
- **Metadata format and file structure**
- **Vocabulary and ontology**
- **Interoperability**
- **Standards development organizations vs. grass roots efforts**
- **General vs. domain-specific**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



GENERATE/ACQUIRE: Acquisition Software Subtopics

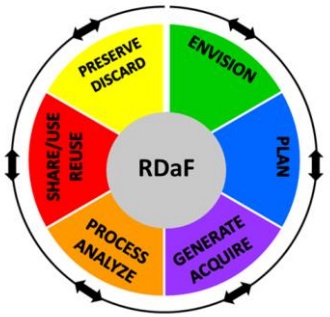
- **Electronic laboratory notebooks**
- **Laboratory information management systems**
- **Instrument control**
- **Open source vs. proprietary**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to GENERATE/ACQUIRE Main Page



PROCESS/ANALYZE Lifecycle Stage

Types of Processed Data

- Tables, spreadsheets
- Charts, graphs
- Maps, vectors, images
- Structured, e.g., hierarchical organization
- Evaluated datasets
- Datasets from models and simulations
- Instrument outputs
- Dynamic data

Scientific Workflow Processes

- Workflow tools
- LIMS
- Laboratory notebooks, e.g., electronic paper
- Containerization
- Reusable workflow components
- Microservices
- Interactive development environments, e.g., Jupyter, Rstudio
- Data modeling and analytics

Data Processing Methods

- Manual
- Automated, autonomous
- Descriptive
- Diagnostic
- Predictive
- Prescriptive

Data Provenance

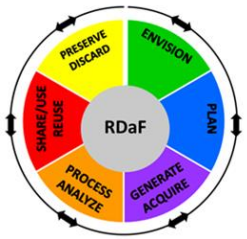
- Original authoritative copy
- Version identification
- Provenance of data derived from other data
- Provenance of scientific records across all the aggregated and differentiated outputs
- Timestamping
- CRediT Taxonomy for all contributors

Metadata

- Types of metadata
- Responsible parties
- Specification of metadata standards
- Linked data structure
- Persistent identifiers (e.g., DOI, ORCID, ARK, ROR, PIDINST, Handles)

Software

- Commercial and/or custom tool
- Open-source vs proprietary
- Capability of tools
- Data aggregation tools
- Surveying tools
- Statistical tools
- Calculation and analysis tools
- Database management tools
- Testing and validation tools
- Documentation
- Reproducibility and uncertainty quantification
- Versioning and maintenance
- Systems resilience and adaptability
- Source code repository



PROCESS/ANALYZE: Topics

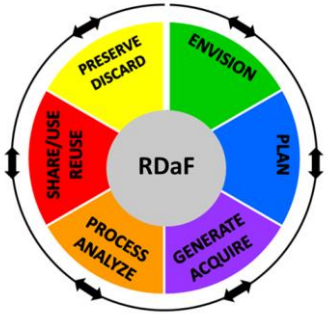
- Types of Processed Data
- Data Processing Methods
- Data Provenance
- Metadata
- Software
- Scientific Workflow Processes

Which topics are relevant to your job role or do you have influence over?

Are some topics more important than others?

Are any topics missing?

Return to PROCESS/ANALYZE Main Page



PROCESS/ANALYZE: Types of Processed Data Subtopics

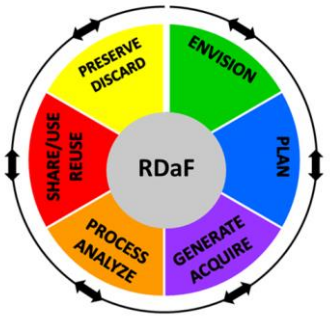
- Tables, spreadsheets
- Charts, graphs
- Maps, vectors, images
- Structured, e.g., hierarchical organization
- Evaluated datasets
- Datasets from models and simulations
- Instrument outputs
- Dynamic data

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PROCESS/ANALYZE Main Page



PROCESS/ANALYZE: Data Processing Methods

Subtopics

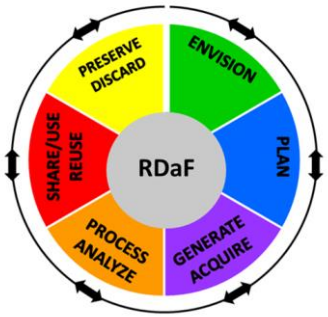
- **Manual**
- **Automated, autonomous**
- **Descriptive**
- **Diagnostic**
- **Predictive**
- **Prescriptive**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PROCESS/ANALYZE Main Page



PROCESS/ANALYZE: Data Provenance Subtopics

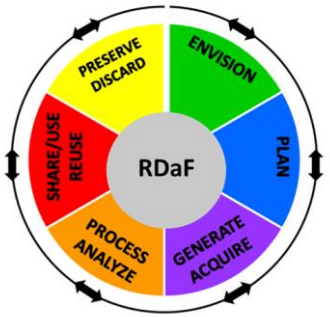
- **Original authoritative copy**
- **Version identification**
- **Provenance of data derived from other data**
- **Provenance of scientific records across all the aggregated and differentiated outputs**
- **Timestamping**
- **CRediT Taxonomy for all contributors**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PROCESS/ANALYZE Main Page



PROCESS/ANALYZE: Metadata Subtopics

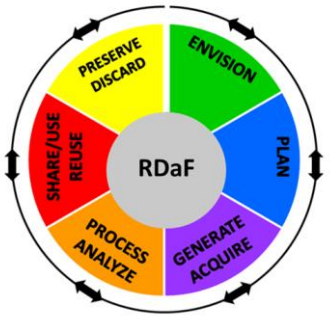
- **Types of metadata**
- **Responsible parties**
- **Specification of metadata standards**
- **Linked data structures**
- **Persistent identifiers (e.g., DOI, ORCID, ARK, ROR, PIDINST, Handles)**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PROCESS/ANALYZE Main Page



PROCESS/ANALYZE: Software Subtopics

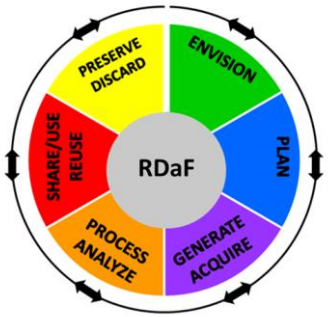
- Commercial and/or custom tools
- Open-source vs proprietary
- Capability of tools
- Data aggregation tools
- Surveying tools
- Statistical tools
- Calculation and analysis tools
- Database management tools
- Testing and validation tools
- Documentation
- Reproducibility and uncertainty quantification
- Versioning and maintenance
- Systems resilience and adaptability
- Source code repository

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PROCESS/ANALYZE Main Page



PROCESS/ANALYZE: Scientific Workflow Processes

Subtopics

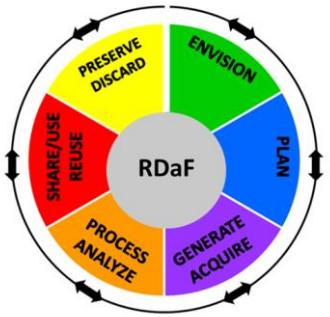
- **Workflow tools**
- **Laboratory information management systems**
- **Laboratory notebooks, e.g., electronic, paper**
- **Containerization**
- **Reusable workflow components**
- **Microservices**
- **Interactive development environments, e.g., Jupyter, RStudio**
- **Data modeling and analytics**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PROCESS/ANALYZE Main Page



SHARE/USE/REUSE Lifecycle Stage

Data Publishing

- Repositories
- Referencing data/digital objects from journal articles
- Supplementary material
- Data linking
- Persistent identifiers
- Metadata
- Quality measures
- Peer review of data sets

Data Attribution

- Citation metrics
- Citation impact
- Citation of combined and derived datasets
- Provenance
- Author identity management
- Use of persistent identifiers
- Data accessibility statement

Data Inventory

- Formats and standards
- Catalog

Internal and External Data Access

- Internal access e.g., data generator
- External access
- Programmatic access aka through Smart API
- Data access vs. data visiting
- Economic constraints

Legal and Licenses

- Ownership of data
- Constraints and encouragement for data use
- Intellectual property rights/restrictions
- Usage agreements/terms/licenses and required permissions
- Terms of Service
- Data sharing agreements and licensing
- Data citation

Levels of Protection

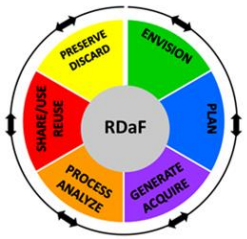
- Unclassified but sensitive information e.g., de-identification, enclaves
- Security classification
- Protecting limited data/secure platforms/enclaves
- Data anonymization

Data Architectures for Application and Use

- Extensibility across communities, including machine-based interactions
- Capturing insights from machine learning and use of these to improve datasets for future artificial intelligence applications
- Capturing data performance characteristics
- Location of data e.g., relative to instruments, in the cloud, transient copies

New Applications and Analyses

- Technologies for use and analytics, e.g., artificial intelligence, machine learning
- Possible repurposing



SHARE/USE/REUSE : Topics

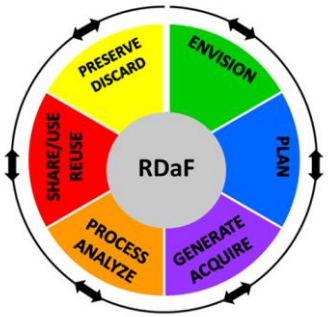
- Data Publishing
- Data Attribution
- Internal and External Data Access
- Legal and Licenses
- Levels of Protection
- Data Architectures for Application and Use
- New Applications and Analyses
- Data Inventory

Which topics are relevant to your job role or do you have influence over?

Are some topics more important than others?

Are any topics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: Data Publishing Subtopics

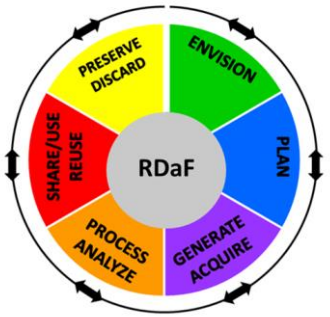
- **Repositories**
- **Referencing data/digital objects from journal articles**
- **Supplementary material**
- **Data linking**
- **Persistent identifiers**
- **Metadata**
- **Quality measures**
- **Peer review of datasets**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: Data Attribution Subtopics

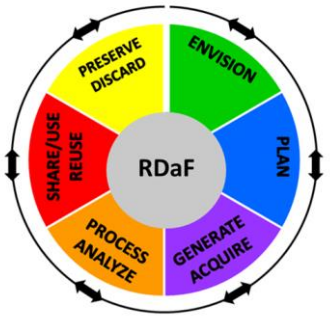
- **Citation metrics**
- **Citation impact**
- **Citation of combined and derived datasets**
- **Provenance**
- **Author identity management**
- **Use of persistent identifiers**
- **Data accessibility statement**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: Internal and External Data Access Subtopics

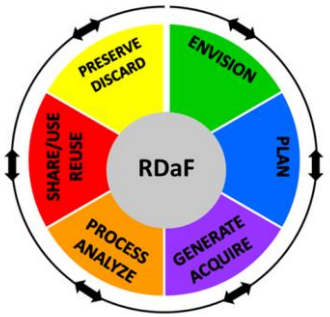
- **Internal access, e.g., data generator**
- **External access**
- **Programmatic access, e.g., through Smart APIs (application programming interfaces)**
- **Data access vs. data visiting**
- **Economic constraints**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: Legal and Licenses Subtopics

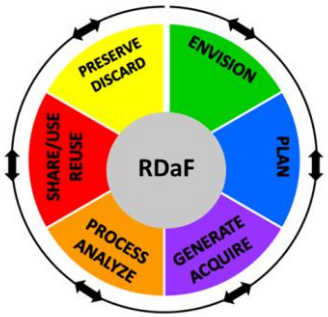
- **Ownership of data**
- **Constraints and encouragement for data use**
- **Intellectual property rights/restrictions**
- **Usage agreements/terms/licenses and required permissions**
- **Terms of Service**
- **Data sharing agreements and licensing**
- **Data citation**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: Levels of Protection Subtopics

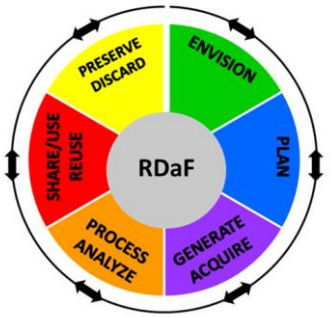
- **Unclassified but sensitive information, e.g. de-identification, enclaves**
- **Security classification**
- **Protecting limited data/secure platforms/enclaves**
- **Data anonymization**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: Data Architectures for Application and Use Subtopics

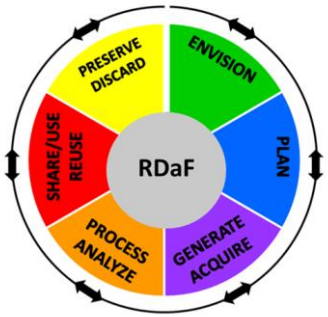
- **Extensibility across communities, including machine-based interactions**
- **Capturing insights from machine learning and use of these to improve datasets for future artificial intelligence applications**
- **Capturing data performance characteristics**
- **Location of data, e.g., relative to instruments, in the cloud, transient copies**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: New Applications and Analyses Subtopics

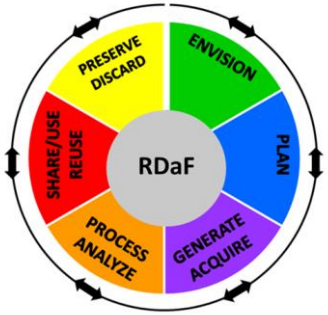
- **Technologies for use and analytics, e.g., artificial intelligence, machine learning**
- **Possible repurposing**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



SHARE/USE/REUSE: Data Inventory Subtopics

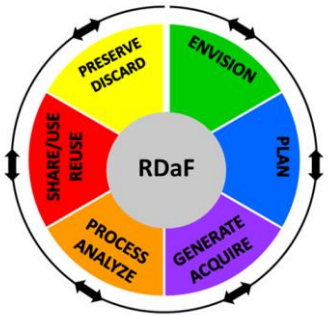
- **Formats and standards**
- **Catalogs**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to SHARE/USE/REUSE Main Page



PRESERVE/DISCARD Lifecycle Stage

Criteria for Preservation

- Use
- Impact
- Value
- Uniqueness
- Cost
- Provenance
- Legal and regulatory

Data Sustainability

- Data longevity and support
- Orphan datasets

Storage and Preservation of Data

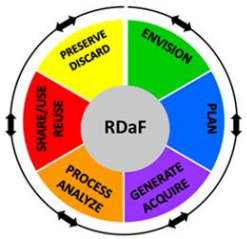
- Media to store and preserve data
- Data back-up
- Data repositories
- Ability to do advanced searches
- File Integrity
- Data Recovery

Moving Data from One Service to Another across Organizations

- Roles and responsibilities
- Moving data from one agency to another
- Registration of repositories: roles and responsibilities
- Disciplinary archives

Retention and Disposition Schedules

- Technical decisions (Data archiving)
- Administrative/policy decisions
- Deaccessioning/End-of-life
- End-of-life special considerations
- Recognition of removed data (tombstone)



PRESERVE/DISCARD: Topics

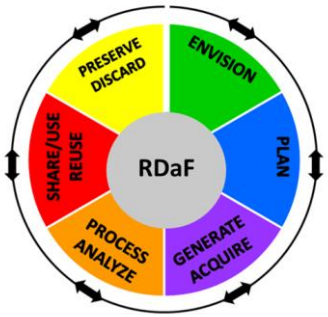
- Criteria for Preservation
- Data Sustainability
- Storage and Preservation of Data
- Moving Data from One Service to Another across Organizations
- Retention and Disposition Schedules

Which topics are relevant to your job role or do you have influence over?

Are some topics more important than others?

Are any topics missing?

Return to PRESERVE/DISCARD Main Page



PRESERVE/DISCARD: Criteria for Preservation Subtopics

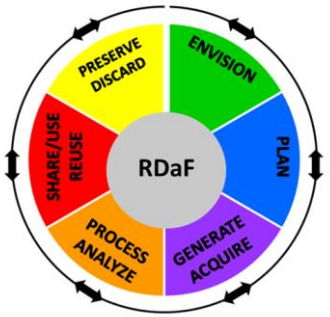
- **Use**
- **Impact**
- **Value**
- **Uniqueness**
- **Cost**
- **Provenance**
- **Legal and regulatory**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PRESERVE/DISCARD Main Page



PRESERVE/DISCARD: Data Sustainability Subtopics

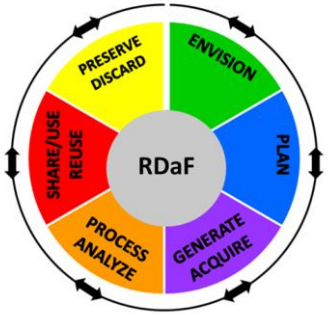
- **Data longevity and support**
- **Orphan datasets**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PRESERVE/DISCARD Main Page



PRESERVE/DISCARD: Storage and Preservation of Data Subtopics

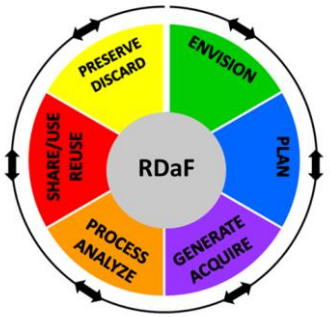
- **Media to store and preserve data**
- **Data back-up**
- **Data repositories**
- **Ability to do advanced searches**
- **File integrity**
- **Data recovery**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PRESERVE/DISCARD Main Page



PRESERVE/DISCARD: Moving Data from One Service to Another across Organizations Subtopics

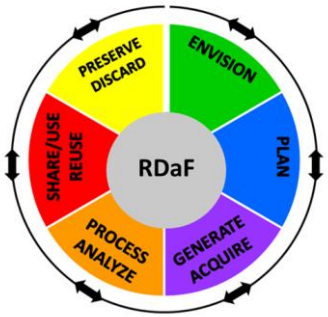
- **Roles and responsibilities**
- **Moving data from one agency to another agency**
- **Registration of repositories: roles and responsibilities**
- **Disciplinary archives**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PRESERVE/DISCARD Main Page



PRESERVE/DISCARD: Retention and Disposition Schedules Subtopics

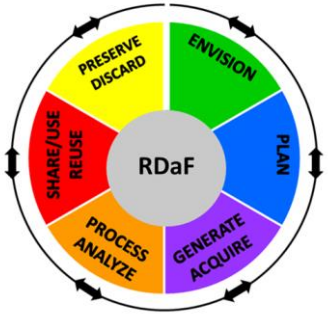
- **Technical decisions (data archiving)**
- **Administrative/policy decisions**
- **Deaccessioning/End-of-life**
- **End-of-life special considerations**
- **Recognition of removed data (tombstone)**

Which subtopics are relevant to your job role or do you have influence over?

Are some subtopics more important than others?

Are any subtopics missing?

Return to PRESERVE/DISCARD Main Page



Next Steps

- Based on feedback from this workshop, the NIST team will develop a generic profile which will serve as a checklist of the subtopics that a person in this role should be aware of in conducting their job
- These generic profiles will serve as starting points for individuals seeking to customize the framework. If you are interested in going through the process of defining a customized profile for yourself, we would welcome that input and any comments you have about the process of doing a personalized profile.
- The NIST team may interview you, other individuals, or small groups to get more input on profiles. If you have comments in the meantime, please contact us at rdaf@nist.gov
- Updates will also be made to the entire framework document with a interim version 1.5 to be released in late 2022/early 2023 for public comment
- Comments may be incorporated in version 2.0 of the framework to be released in spring 2023

Thanks for your invaluable participation!