

## RDaF Preliminary Framework Core

(Appendix E in NIST publication [SP 1500-18](#))

**Notes:**

- 1) In the Categories and Subcategories, “data” means “research data;”
- 2) Bolded words indicate input from the Stakeholder Scoping Workshop; and
- 3) A \* at the end of a word or group of words indicates that a definition is provided in Appendix G of SP 1500-18)

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
<p><b>ENVISION</b></p> <p>Review of the overall strategies and drivers of an organization’s research data program.</p>	Data <b>Governance*</b> Structure	<ul style="list-style-type: none"> <li>• Identification of Goals and Roles</li> <li>• Data vision and/or <b>data policy</b></li> <li>• Data management value proposition</li> <li>• Data management organization</li> <li>• Value of data (quantitative or qualitative)</li> <li>• Legal and regulatory <b>compliance</b></li> <li>• <b>Data quality</b> (including Trust and Certification)</li> <li>• <b>Data privacy</b></li> <li>• Data ethics</li> </ul>
	<b>Community</b> Engagement	<ul style="list-style-type: none"> <li>• <b>Stakeholder community(ies)</b></li> <li>• Communication with stakeholder community(ies)</li> <li>• Interactions with other organizations</li> <li>• <b>Cross-community engagement</b> (across domains and sectors)</li> <li>• Inclusivity in interactions</li> </ul>
	<b>Data Culture*</b>	<ul style="list-style-type: none"> <li>• <b>FAIR data principles</b></li> <li>• <b>Value of data</b></li> <li>• Roles and responsibilities</li> </ul>
	<b>Reward Structure</b>	<ul style="list-style-type: none"> <li>• For data management</li> <li>• Value of <b>data workers</b></li> <li>• <b>Incentives and institutional credit for data sharing</b> and reuse</li> <li>• Disincentives for data sharing</li> <li>• Human Resources (HR) involvement</li> </ul>

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
<b>ENVISION</b> (continued)	<b>Workforce/Career Paths</b>	<ul style="list-style-type: none"> <li>• Workforce skills inventory</li> <li>• HR's role in data <b>workforce development</b></li> <li>• Data management <b>training</b></li> <li>• Workforce preparedness in <b>new and advancing technologies</b>, e.g., HPC, AI, ML, and computation services</li> <li>• Promotional paths, continual training, and career development</li> </ul>
	<b>Data Safety and Security</b>	<ul style="list-style-type: none"> <li>• Safety and security assurance</li> <li>• Data inventory</li> </ul>
	Strategy	<ul style="list-style-type: none"> <li>• Organizational data management</li> </ul>
	<b>Data Risk Management*</b>	<ul style="list-style-type: none"> <li>• Risk assessment</li> <li>• Risk mitigation and management</li> </ul>
<b>PLAN</b> The tactical management positioning in an organization for effective research data management throughout the research data lifecycle*.	Chain of Control	<ul style="list-style-type: none"> <li>• Documentation</li> <li>• Communication within organization</li> </ul>
	<b>Economics and Costs of Planning</b>	<ul style="list-style-type: none"> <li>• Decision-making tools for data, including cost-benefit analysis</li> <li>• Cost breakdown, i.e., calculation of costs by data lifecycle* stage</li> </ul>
	Funding Planning	<ul style="list-style-type: none"> <li>• Models for <b>provisioning resources</b>, i.e., direct, overhead, or mixed</li> </ul>
	Data Objects	<ul style="list-style-type: none"> <li>• Quantitative and qualitative data</li> <li>• Software, models</li> <li>• Instruments</li> <li>• Data publications*, journal publications</li> <li>• Presentations</li> <li>• Other</li> </ul>
	Hardware/Software Infrastructure	<ul style="list-style-type: none"> <li>• Interoperability</li> <li>• Persistent instrument identifiers</li> </ul>
	Data Management Planning	<ul style="list-style-type: none"> <li>• <b>Data management plans (DMPs)</b></li> <li>• Lifecycle considerations: living documents or static proposals?</li> </ul>
	<b>Scientific Data Standards</b>	<ul style="list-style-type: none"> <li>• Sources of standards</li> <li>• General, domain-specific</li> </ul>
	Assessment and Controls	<ul style="list-style-type: none"> <li>• Goals/definition of success</li> <li>• <b>Metrics</b> or metrics structure, tracking use and impact measures</li> </ul>

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
<p><b>GENERATE/ ACQUIRE</b></p> <p>The generation of raw research data and/or the acquisition* of research data by an organization.</p>	Sources of Raw Data*	<ul style="list-style-type: none"> <li>Generated In-house experimentally or computationally</li> <li>Collected from external sources</li> </ul>
	Experimental Data Generation	<ul style="list-style-type: none"> <li>Specification and recording of instruments and associated metadata</li> <li>Description and recording of measurement protocols</li> <li>Methods for data and metadata capture and recording</li> </ul>
	Computational Data Generation	<ul style="list-style-type: none"> <li>Commercial and/or custom software</li> <li>Methods for computational variables (metadata) capture and recording</li> </ul>
	FAIR Principles for Data Generated In-House	<ul style="list-style-type: none"> <li>Data born FAIR</li> <li>Data made FAIR</li> </ul>
	External Sources of Data	<ul style="list-style-type: none"> <li>Data acquired FAIR</li> <li><b>Identification, collection, and recording</b></li> <li><b>Metadata harvesting</b></li> </ul>
	<b>Community-Based Standards for Formats</b>	<ul style="list-style-type: none"> <li>Standards development organizations/sources</li> <li>General, domain-specific</li> </ul>
<p><b>PROCESS/ ANALYZE</b></p> <p>The actions performed on generated or acquired research data to yield processed research data, and the research data stewardship* functions performed by an organization.</p>	Data Provenance	<ul style="list-style-type: none"> <li>Original authoritative copy</li> <li>Version identification</li> <li>Provenance of data derived from other data</li> <li>Provenance of scientific records across all the individual outputs</li> <li><b>Timestamping</b></li> </ul>
	Data Architecture	<ul style="list-style-type: none"> <li>Design</li> <li>Security</li> <li>Configuration management</li> <li>Hosting and storage</li> <li>Use of cloud</li> </ul>
	<b>Software Tools</b>	<ul style="list-style-type: none"> <li>Data lifecycle*</li> <li>Management and analysis</li> <li>Commercial and/or custom tools</li> <li>System resilience and adaptability</li> <li>Maintenance</li> </ul>

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
<b>PROCESS/ ANALYZE</b> (continued)	Scientific <b>Workflow</b> Processes and Systems	<ul style="list-style-type: none"> <li>• Workflow tools</li> <li>• Laboratory notebooks, i.e., electronic, paper</li> </ul>
	Data Inventory	<ul style="list-style-type: none"> <li>• Formats and standards</li> <li>• Catalogs</li> <li>• Interoperability (across instrument manufacturer file formats)</li> </ul>
	Data Modeling and Analytics	<ul style="list-style-type: none"> <li>• Processes</li> <li>• Tools</li> </ul>
	<b>Data Representation/ Models/Structures</b>	<ul style="list-style-type: none"> <li>• Dynamic data</li> <li>• General, domain-specific</li> </ul>
	<b>Data Curation</b>	<ul style="list-style-type: none"> <li>• Policies and processes</li> <li>• Manpower</li> </ul>
	<b>Metadata</b>	<ul style="list-style-type: none"> <li>• Types of metadata</li> <li>• Responsible parties</li> <li>• Specification of metadata standards</li> <li>• <b>Linked data</b> structure</li> <li>• Persistent identification (DOI)</li> </ul>
<b>SHARE/USE/ REUSE</b>  How research data are disseminated, used, and reused within and outside an organization.	Legal and Licenses	<ul style="list-style-type: none"> <li>• Ownership of data</li> <li>• Constraints and encouragement for data use</li> <li>• <b>Intellectual property</b> rights/restrictions</li> <li>• <b>Usage agreements/terms/licenses</b> and required permissions</li> <li>• Terms of service</li> <li>• Data sharing agreements and licensing</li> <li>• Data citation*</li> </ul>
	<b>Data Publishing*</b>	<ul style="list-style-type: none"> <li>• Repositories</li> <li>• Referencing data/digital objects from journal articles</li> <li>• Supplementary material</li> <li>• Data linking</li> </ul>
	Data Citation*	<ul style="list-style-type: none"> <li>• Citation metrics</li> <li>• Citation impact</li> </ul>
	Internal and External Data Access	<ul style="list-style-type: none"> <li>• Access internally, e.g., the data generator</li> <li>• Access externally</li> <li>• <b>Programmatic access</b>, aka Smart API</li> <li>• Data access vs. data visiting</li> </ul>
	Levels of Protection	<ul style="list-style-type: none"> <li>• Unclassified but sensitive information, e.g., de-identification, enclaves</li> <li>• Security classification</li> </ul>

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
	Levels of Protection (continued)	<ul style="list-style-type: none"> <li>Protecting limited data/secure platforms/enclaves</li> <li>Data anonymization*</li> </ul>
	Applications and Analysis	<ul style="list-style-type: none"> <li><b>Technologies</b> for use and analytics, e.g., AI, ML</li> </ul>
	Data Architectures for <b>Application and Use</b>	<ul style="list-style-type: none"> <li>Extensibility across communities, including machine-based interactions</li> <li>Capturing insights from ML and use of these to improve datasets for future AI applications</li> <li>Capturing data performance characteristics</li> <li>Location of data (e.g., relative to instruments, in the cloud, transient copies)</li> </ul>
<p><b>PRESERVE/ DISCARD</b></p> <p>The end-of-use and end-of-life provisions for research data in an organization, including records management, archiving, and safe disposal.</p>	Criteria	<ul style="list-style-type: none"> <li>Use and impact</li> </ul>
	<b>Data Sustainability</b>	<ul style="list-style-type: none"> <li>Data longevity and support</li> <li>Orphan datasets</li> </ul>
	<b>Storage and Preservation of Data</b>	<ul style="list-style-type: none"> <li>Media to store and preserve data</li> <li>Data back-up</li> <li>Data repositories</li> </ul>
	Moving Data from One Service to Another across Organizations	<ul style="list-style-type: none"> <li>Roles and responsibilities</li> <li>Moving data from one agency to another, e.g., from a funded research agency to an agency with a permanent repository</li> <li>Registration of repositories: roles and responsibilities</li> <li>Disciplinary archives</li> </ul>
	<b>Retention and Disposition Schedules</b>	<ul style="list-style-type: none"> <li>Data <b>archiving</b>, i.e., <b>what is kept and not kept</b> <ul style="list-style-type: none"> <li>Decision processes</li> </ul> </li> <li><b>End-of-life</b> issues <ul style="list-style-type: none"> <li>Example: Responsible party for keeping raw data* feeds</li> <li>Example: Store (or not) raw data*, given the large amount of storage needed</li> </ul> </li> <li>Deaccessioning/End-of-life</li> <li>Recognition of removed data (gravestone)</li> </ul>