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

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

NIST Special Publication 1500-18

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

Robert J. Hanisch
Debra L. Kaiser
Office of Data and Informatics
Material Measurement Laboratory

Bonnie C. Carroll
Consultant
Information International Associates
Oak Ridge, TN

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

February 2021



U.S. Department of Commerce
Wynn Coggins, Acting Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Publications in the SP1500 subseries are intended to capture external perspectives related to NIST standards, measurement, and testing-related efforts. These external perspectives can come from industry, academia, government, and others. These reports are intended to document external perspectives and do not represent official NIST positions. The opinions, recommendations, findings, and conclusions in this publication do not necessarily reflect the views or policies of NIST or the United States Government.

National Institute of Standards and Technology Special Publication 1500-18
Natl. Inst. Stand. Technol. Spec. Publ. 1500-18, 42 pages (February 2021)
CODEN: NSPUE2

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

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

Foreword

This publication is the result of an ongoing collaborative effort involving industry, government agencies, universities, non-profits, publishers, and other organizations and institutions. The National Institute of Standards and Technology (NIST) launched the Research Data Framework (RDaF) project by convening national and international private- and public-sector organizations and individuals at a workshop in December 2019. Attendees at this workshop unanimously agreed that the RDaF should move forward and that NIST was the best institution to take the lead in its development. Further, it was recommended that NIST move as rapidly as possible to solidify the plan and seek collaborative funding with other government agencies.

Abstract

NIST is leading the development of the Research Data Framework (RDaF) with involvement and input from national and international leaders in the broad research data stakeholder community. Research data is defined here as "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings." The overarching goal of the RDaF is to provide the stakeholder community with a structured approach to develop a customizable strategy for the management of research data. The audience for the RDaF is the entire research data community, including all organizations and individuals engaged in any activities concerned with research data management, from Chief Executive Officers and Chief Data Officers to librarians and researchers. This document describes the motivation for, and the development of, a Preliminary Framework Core, and identifies the next steps in further development of the RDaF. The research data environment is rapidly changing, and this Framework shall remain a living document. Revisions will be made as we, the stakeholders of the RDaF, gain experience with its application and use.

Key words

Research data; research data ecosystem; research data framework; research data lifecycle; research data management; stakeholder community.

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

Table of Contents

1. Introduction 1

1.1. Motivation 1
1.2. Origin of the Framework 2
1.3. What is the RDaF? 3
1.4. Legal and Institutional Drivers 3
1.5. Value Proposition 4
1.6. Risk Management 5

2. Development of the Preliminary RDaF 6

2.1. Initial Scoping Study 6
2.2. Stakeholder Scoping Workshop 7
2.3. Interim Studies and Reports 8
2.4. Drafting the Preliminary RDaF 8

3. Description of the Preliminary RDaF 8

3.1. Relationship to Other NIST Frameworks 8
3.2. Framework Core 9
3.3. Informative References 11
3.4. Framework Profiles 11
3.5. Framework Implementation Tiers 12

4. Next Steps 12

Acknowledgments 13

References 14

Appendix A: Acronyms and Initialisms 18

Appendix B: Initial List of Stakeholders and Users 20

Appendix C: RDaF Stakeholder Scoping Workshop Agenda 23

Appendix D: RDaF Stakeholder Scoping Workshop Attendees 24

Appendix E: Preliminary RDaF Framework Core 27

Appendix F: Initial List of Informative References 32

Appendix G: Glossary of Terms used in Appendix E 34

List of Tables

Table 1. RDaF Steering Committee members. 5
Table 2. Core Functions of the NIST Frameworks. 9
Table 3. Example of Framework Profile development. 12

List of Figures

Fig. 1. Timeline for development of the Preliminary RDaF..... 6
Fig. 2. Four elements of a Framework Core. 9
Fig. 3. The six Functions (research data lifecycle stages) of the Preliminary RDaF 11
Fig. 4. Timeline for the two pilot studies..... 13

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

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

1. Introduction

NIST is leading the development of the Research Data Framework (RDaF) with involvement and input from national and international leaders in the broad research data stakeholder community. Research data is defined here as "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings [1]." The overarching goal of the RDaF is to provide the stakeholder community with a structured approach to develop a customizable strategy for the management of research data. The audience for the RDaF is the entire research data community, including all organizations and individuals engaged in any activities concerning research data management, from Chief Executive Officers (CEOs) and Chief Data Officers (CDOs) to librarians and researchers. This document is organized into four high-level Sections: (1) Introductory material; (2) Development of an initial version of the RDaF, hereafter referred to as the Preliminary RDaF, (3) A description of the Preliminary RDaF; and (4) Next steps.

1.1. Motivation

It is widely recognized that data, specifically research data, are of growing importance and impact to the economy and society.

> *"Data and information have become the most important assets of the 21st century* [2]*."*
>
> "*Data-driven innovation forms a key pillar in 21st century sources of growth…[research] data sets are becoming a core asset in the economy, fostering new industries, processes and products…[such data sets] significantly enhance productivity, resource efficiency, economic competitiveness, and social well-being* [3]."
>
> *"Scientific research supported by the [U.S.] federal government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security* [4]*."*
>
> *"We are living in a data explosion where we generate and consume [research] data faster than we can keep track of and secure. What are we going to do with all this [research] data and how can we unlock its potential to make it work for society?* [5]*"*

The risks of losing and mismanaging research data can have severe economic and social consequences. [6–9]. Europe and China have recognized this and have moved proactively to develop federated enterprise approaches to manage research data and make such data widely available. Europe has taken the leadership position in open research with FAIR (Findable, Accessible, Interoperable, Reusable) [10] data and is moving toward implementation of FAIR with the European Open Science Cloud [11]. China is also working aggressively on open research with its China Science and Technology Cloud started in October 2020 [12]. More recently, the Chinese Academy of Sciences has funded a five-year international partnership pilot project, the Global Open Science Cloud Initiative, to begin in early 2021 [13]. The U.S. has also recognized the criticality of open science through a National

Academies[1] study, *Open Science by Design: Realizing a Vision for 21st Century Research* [14].

For decades, advances in information technology offer unprecedented opportunities in scientific data [15]. Consider, for example, supercomputers that can perform a million trillion total operations per second [16]. The U.S. has the second fastest scientific research computer in the world (IBM Summit), having been recently displaced from the leading position in June 2020 by the Japanese supercomputer Fugaku [17]. U.S. researchers are applying Summit to key global issues, including artificial intelligence (AI) [18] and genetic decoding of COVID-19 [19]. The U.S. has many other active private, academic, and public investments in data-intensive research in myriad disciplines.

There is an increasing variety of stakeholders in the research data ecosystem: government agencies (e.g., the OSTP/NSTC Subcommittee on Open Science[2] in the U.S.), universities and their research libraries, research data centers and repositories, scholarly publishers, professional societies, national and international collaborations, organizations (e.g., CENDI,[3] BRDI,[4] NASEM,[5] CODATA,[6] RDA,[7] WDS,[8] and GO FAIR,[9] see Appendix A, Acronyms and Initialisms), standards bodies, funders (both public and private), industry and the private sector, researchers, and the general public. How do the roles, responsibilities, and expectations of these diverse stakeholders differ, overlap, or contradict?

1.2. Origin of the Framework

The concept of a Research Data Framework (RDaF) is inspired by the demonstrated success of the *Framework for Improving Critical Infrastructure Cybersecurity* [20], which NIST initially issued in February 2014, and which is hereafter referred to as the NIST Cybersecurity Framework.

The development of the RDaF started with a preliminary scoping study to determine the best approach to get support and uptake from a diverse stakeholder community. The RDaF will focus on the U.S., but by necessity will include global players and global best practices. Open and FAIR data are essential tenets in the Framework, but it supports the concept of "as open as possible, as closed as necessary [21, 22]." The details of this initial version of the RDaF, hereafter termed the Preliminary RDaF, were informed by a small but representative subset of the research data community. Subsequent versions of the RDaF will be informed by the broader community.

[1] The National Academies of Sciences, Engineering, and Medicine. Available at https://www.nationalacademies.org/

[2] Office of Science and Technology Policy, National Subcommittee on Science and Technology. Available at https://web.archive.org/web/20201201153151/https://www.whitehouse.gov/ostp/nstc/ under Committee on Science.

[3] Available at https://web.archive.org/web/20201109215414/https://www.cendi.gov/

[4] Available at https://www.nationalacademies.org/brdi/board-on-research-data-and-information

[5] Available at https://www.nationalacademies.org/home

[6] Available at https://web.archive.org/web/20201207173554/https://codata.org/

[7] Available at https://web.archive.org/web/20201207174444/https://rd-alliance.org

[8] Available at https://web.archive.org/web/20201207174825/https://www.worlddatasystem.org/

[9] Available at https://web.archive.org/web/20201101081800/https://www.go-fair.org/

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

The research data space is replete with well-intentioned and useful initiatives. However, these initiatives, which involve only one or a few of the stakeholder groups listed in Section 2.1, cannot be representative of the entire research data ecosystem. There are integrated efforts within CODATA[4] and RDA,[5] and topical programs such as the Materials Genome Initiative,[10] the Global Biodiversity Information Facility,[11] and the BRAIN Initiative.[12] The RDaF will take advantage of this plethora of activities and organizations to facilitate better coordination and thus assure maximum return on the investment in research data infrastructure and interoperability tools.

1.3. What is the RDaF?

The research data ecosystem is very complex! There are many stakeholders and various funding models and sustainability plans. How long should research data be kept? How should research data quality be assessed? How do we measure the value of research data? The RDaF strives to answer these questions by providing:

- A map of the research data space: who, what, where, why, and 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 among the stakeholders in research data; and
- A tool that may be used to change the research data culture in an organization.

1.4. Legal and Institutional Drivers

The RDaF aspires to provide organizations with a structured approach to develop a coherent research data strategy, and to provide stakeholders with some common language terms[13] and a basis for coordination. NIST will lead the coordinated effort to develop and maintain a Framework that is useful but voluntary for all sectors of the economy, e.g., industry, government, academia, and not-for-profit organizations.

Just as the first version of the NIST Cybersecurity Framework was initially driven by legislation, namely Executive Order 13636: *Improving Critical Infrastructure Cybersecurity* [23], there are federal directives that support the development of the RDaF. These include a series of White House directives, with the most influential being *Increasing Access to the Results of Federally Funded Scientific Research* [4], also known as "the Holdren memo," which was issued in February 2013. This memorandum was followed by another memorandum, *Open Data Policy-Managing Information as an Asset* in May 2013, and by Executive Order 13642: *Making Open and Machine Readable the New Default for Government Information* (May 14, 2013) [24]." On January 14, 2019, President Obama signed into law the *Foundations for Evidence-Based Policymaking Act* of 2018 [25], which includes the *OPEN Government Data Act*, House Resolution 1770 [26]. This legislation collectively dictates that U.S, government agencies must make their data publicly available. Complying with these national requirements and taking into consideration the massive efforts

[10] Available at https://web.archive.org/web/20201207180439/https://www.mgi.gov/
[11] Available at https://web.archive.org/web/20201207180743/https://www.gbif.org/
[12] Available at https://web.archive.org/web/20201126165836/https://braininitiative.nih.gov/
[13] Some of the language terms will be dependent on the specific research discipline.

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

in the open research/open data world, the U.S. is assessing and promoting the best practices that are emerging in a diverse and complex global ecosystem of research data [27]. The European Commission, through its European Open Science Cloud [11], aims to create a European research interoperability framework. RDaF leadership will keep abreast of this and other international efforts to achieve a consistent approach across the entire research data lifecycle.

1.5. Value Proposition

The immense value of managing research data is clearly supported by several federal documents. As stated in *Open Data Policy–Managing Information as an Asset* [24],

> *"Managing government information as an asset will increase operational efficiencies, reduce costs, improve services, support mission needs, safeguard personal information, and increase public access to valuable government information. Making information resources accessible, discoverable, and usable by the public can help fuel entrepreneurship, innovation, and scientific discovery—all of which improve Americans' lives and contribute significantly to job creation."*

From 2017 to 2019, the U.S. government released three key documents concerning Federal data: (1) *The Promise of Evidence-Based Policymaking*, which describes improvements on how data are used to generate evidence about policies and programs in the federal government [28]; (2*) The President's Management Agenda: Modernizing Government for the 21st Century*, which sets a priority goal of leveraging data as a strategic asset [29]; and (3) *The President's Management Agenda: Federal Data Strategy 2020 Action Plan* [30], which defines the steps to achieve this goal. The *Foundations for Evidence-Based Policymaking Act of 2018* [25] stipulates the reporting structure for data management as follows:

> *"[To] improve Federal data management…The head of each agency shall designate a nonpolitical appointee employee in the agency as the Chief Data Officer of the agency [who] shall be responsible for lifecycle data management…There is established in the Office of Management and Budget a Chief Data Officer Council that shall (1) establish Government wide best practices for the use, protection, dissemination, and generation of data; [and] (2) promote and encourage data sharing agreements between agencies."*

In July 2019, a Steering Committee consisting of eight individuals from different parts of the research data ecosystem was recruited to assist and advise in the development of the RDaF. The RDaF Steering Committee members, listed in Table 1, established a value proposition for the RDaF to include the following benefits:

- ***Research Integrity***: The RDaF will enable higher-quality, reproducible, and better-characterized research data, and transparency of the research process.
- ***Costs and Efficiency***: The RDaF will aid in establishing and applying best practices to research data management to maximize efficiency and control costs.

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

- ***Risk Management and Reduction***: While risk management and reduction practices are designed to decrease potential negative impacts, they may inadvertently result in missed opportunities. The RDaF will help organizations to assess their current risk positions and to create their own roadmap for improvement, including the management and reduction of risk in business decisions.
- ***Scientific Discovery and Innovation***: Scientific discovery and innovation are critical to global competitiveness. The RDaF will embrace the FAIR principles, which promise to increase scientific productivity through better use and reuse of research data.
- ***Policy Compliance***: The RDaF will assist organizations in attaining compliance with research data management and sharing policies from funding organizations and journals/publishers.

Table 1. RDaF Steering Committee members.

Name[a]	Organization	Sector
Laura Biven[b]	Department of Energy	Government
Mercè Crosas	Harvard University	Academia
Joshua Greenberg	Sloan Foundation	Funder, private foundation
Hilary Hanahoe	Research Data Alliance (RDA)	International data organization
Heather Joseph	Scholarly Publishing and Academic Resources Coalition	A non-government advocacy organization, libraries
Barend Mons	Leiden Univ., CODATA, GO-FAIR	Academia, international data organization
Beth Plale[c]	National Science Foundation	Government, funder
Anita de Waard	Elsevier	Scholarly publisher, private sector

[a]Mark Leggott, Research Data Canada (government), was added to the Committee in mid-2020.
[b]Now at National Institutes of Health (NIH).
[c]Now at Indiana University.

1.6. Risk Management

As stated in the NIST Privacy Framework [31], "risk management is a cross-organizational set of processes that helps organizations to understand how their systems, products, and services may create problems for individuals or the organization and how to develop effective solutions to manage such problems…risk assessments produce the information that can help organizations weigh the benefits of data processing[14] against the risks and determine the appropriate response—sometimes referred to as proportionality." Further, the NIST Privacy Framework demonstrates an application of risk management to data and privacy, whereby an organization "optimizes beneficial uses of data while minimizing adverse consequences for individuals' privacy and society as a whole [31]." The applicability of similar risk management and assessment processes for research data will be considered in the next version of the RDaF.

[14]data processing is a collective set of data actions which include, but are not limited to, collection, retention, logging, generation, transformation, use, disclosure, sharing, transmission, and disposal.

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

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

2. Development of the Preliminary RDaF

Because a framework is only successful if it has buy-in and acceptance from the community, it is important to ensure that a wide range of voices are heard. For research data, the community includes business, academia, government, and other types of stakeholders. It involves roles and players that represent all stages of the research data lifecycle. As noted, the RDaF should be global in scope and reach because the nature and applications of research data are intended for broad adoption. The Preliminary RDaF development process is depicted in the timeline in Fig. 1. The "Community of Interest" (COI) includes the Workshop attendees and others who have expressed interest in following the progress of the RDaF.

2.1. Initial Scoping Study

As a necessary first step, initial research was conducted to characterize the current research data landscape, including:

- Stakeholders and users (see Appendix B);
- Standards and tools already produced and in use;
- Maturity models and indicators (i.e., mechanisms to assess the extent of and success of research data management in an organization); and
- Requirements and gaps in knowledge of best practices, including research data infrastructure.

The preliminary scoping study, conducted in consultation with the RDaF Steering Committee, gauged stakeholder interest and determined the best approach to creating a framework that would have support from and adoption by a diverse stakeholder community.

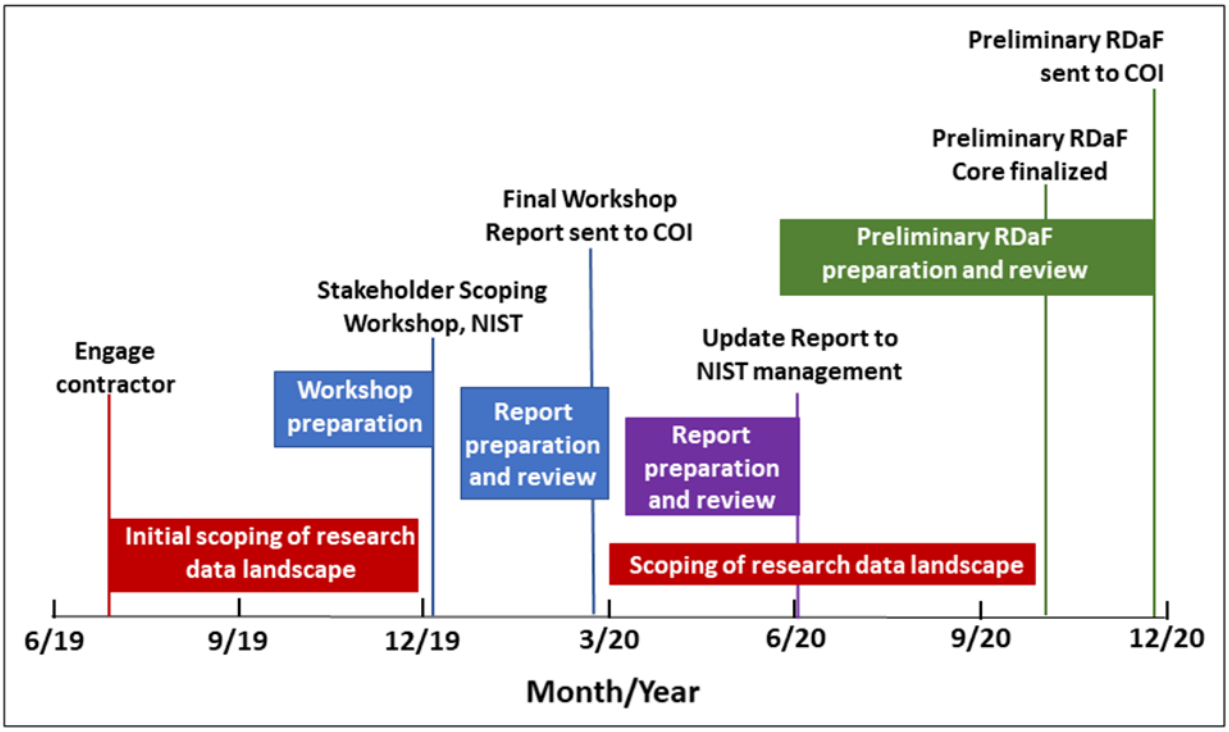


Fig. 1. Timeline for development of the Preliminary RDaF.

2.2. Stakeholder Scoping Workshop

To determine the viability and true value of the RDaF as perceived by the community, an invitation-only Stakeholder Scoping Workshop was held on December 5-6, 2019 at the NIST National Cybersecurity Center of Excellence in Gaithersburg Maryland (see Appendix C for the agenda). The co-chairs of the workshop were Robert Hanisch from NIST and Bonnie Carroll from Information International Associates and CODATA. The RDaF Steering Committee assisted with the technical aspects of the workshop, e.g., identifying attendees and planning the agenda. At the workshop, 50 invited attendees represented a broad spectrum of stakeholders encompassing a variety of job functions within the research data ecosystem. The complete list of workshop attendees is provided in Appendix D. The number distribution of attendees' affiliations is as follows:

- Fifteen from six government departments/agencies
- Eleven from ten U.S. universities
- Five from four U.S. National Laboratories
- Seven from six countries
- Seven from companies in six technology sectors
- Five from non-profit organizations.

All participants actively and enthusiastically engaged in break-out sessions and full-group discussions. The structure of the NIST Cybersecurity Framework resonated with the workshop participants and they recommended its basic structure for the RDaF. Two organizing concepts for the Framework Core were considered at the Workshop: a research data ecosystem approach (where the emphasis is on the interactions between organizations and research data management functions) or a lifecycle approach, perhaps including a top-level "sphere of responsibility." It was decided to use a hybrid approach which is serial in nature and has dynamic processes concerning research data management. Each of four break-out groups proposed various lifecycle stages for the co-chairs and RDaF Steering Committee members to consider in their post-workshop deliberations.

There was consensus that *better national and international coordination is needed now for both basic and applied research data to ensure that the U.S. stays competitive and thinks strategically about the management of such data, a valuable national resource*. All participants were enthusiastic about remaining involved in the RDaF development and it was unanimously agreed that the RDaF should move forward. The following recommendations were made:

- NIST is the best organization to lead the development of the RDaF;
- NIST should move as rapidly as possible to solidify the plan and seek funding from other government agencies and organizations;
- The main target user for the RDaF is at an institutional or organizational level such as a CDO, i.e., someone with broad responsibilities for the management of research data across an organization;
- The RDaF should have value for other roles (i.e., job functions) in organizations, such as researchers;
- Regular communication with the COI should continue by various means, e.g., e-mail updates, webinars; and
- The Steering Committee should be consulted frequently as the RDaF is developed.

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

In summary, the workshop was effective in building the base for moving ahead and for soliciting support for the RDaF development.

In the three months following the workshop, the co-chairs drafted a report which was vetted by the RDaF Steering Committee. This workshop report, *Initial Scoping Study for a NIST-Led Research Data Framework (RDaF)*, was distributed on March 5, 2020 to the RDaF COI. The report contained an initial Framework Core with seven Functions (research data lifecycle stages) and 44 Categories and Subcategories (relevant topics for the seven Functions).

2.3. Interim Studies and Reports

From March 2020 to June 2020, two additional reports were generated. The first was a brief roadmap document, parts of which are incorporated in the present document. The second was a briefing report for NIST upper management and included a budget for continuation of the RDaF project beyond the completion of the initial version presented herein. Scoping of the current research data landscape continued in the four bulleted areas given in Section 2.1 and was used to refine the initial Framework Core mentioned in Section 2.2.

2.4. Drafting the Preliminary RDaF

In the seven months following the distribution of the workshop report, the Framework Core was modified with input from the RDaF Steering Committee. The result was a regrouping of the initial seven Functions to form six Functions. (See Section 3.2, Framework Core.) A detailed description of the Preliminary RDaF is presented in the following Section. A draft version of the present document was vetted by the Steering Committee and released to the RDaF COI on October 26, 2020.

3. Description of the Preliminary RDaF

3.1. Relationship to Other NIST Frameworks

As stated in Section 2.2, a consensus decision was made to base the RDaF structure on that of the successful NIST Cybersecurity Framework, which NIST initially issued in February 2014 to address the similarly emerging and complex global challenge of cybersecurity. Both the NIST Cybersecurity and the NIST Privacy Frameworks have three basic parts: a Framework Core, Framework Profiles, and Framework Implementation Tiers. In these two Frameworks, a Framework Core consists of four elements: Core Functions (activities), Categories and Subcategories (outcomes), and Informative References (e.g., standards, guidelines, and practices). The RDaF Preliminary Framework Core described in the following Section differs from the NIST Cybersecurity and Privacy Framework Cores in one important way: for Categories and Subcategories, *topics* replace *outcomes*. The other two basic parts of the Cybersecurity and Privacy Frameworks, i.e., Framework Profiles and Framework Implementation Tiers, are mentioned briefly in Sections 3.4 and 3.5 in relation to this Preliminary RDaF. These two parts will be included in a future version of the full RDaF.

Completed in October 2019, the nine-volume NIST Big Data Interoperability Framework [32] does not have the three basic parts of the other two completed NIST Frameworks. Future versions of the RDaF will draw upon the Big Data Interoperability Framework as appropriate.

Table 2 compares the Framework Core Functions of the NIST Cybersecurity and Privacy Frameworks with the Core Functions selected for the Preliminary RDaF. (See Section 3.2 for descriptions of the RDaF Core Functions.) The intersections of these three frameworks are

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

evident. For example, Plan, Detect, and Identify all relate to situational awareness; Process/Analyze, Respond, and Control all relate to operational aspects; and Preserve/Discard, Recover, and Protect all relate to final actions.

Table 2. Core Functions of the NIST Frameworks.

RDaF	Cybersecurity	Privacy
Envision Plan Generate/Acquire Process/Analyze Share/Use/Reuse Preserve/Discard	Identify Protect Detect Respond Recover	Identify Govern Control Communicate Protect

3.2. Framework Core

To date, collaborative development of the Preliminary RDaF has focused solely on the Framework Core. The relationship between the four different elements of the Framework Core—Functions, Categories, Subcategories, and Informative References—is shown in Fig. 2 for a Core with three Functions. Definitions of the four elements in the context of research data for the RDaF are as follows:

(1) **Functions** organize foundational research data-related activities at their highest level. As stated in Section 2.2, a hybrid research data lifecycle–ecosystem approach was selected as the organizing concept of the Framework Core.
(2) **Categories** are topics for a Core Function that are closely tied to programmatic needs and activities, as well as other important factors.
(3) **Subcategories** further divide a Category into more specific topics.
(4) **Informative References** are standards, guidelines, and practices associated with a Subcategory that provide the means to address that topic. Informative References will likely be a combination of resources that are common to all disciplines, organizations, and roles as well as resources that are specific to the disciplines, organizations, and roles to which the RDaF is being applied.

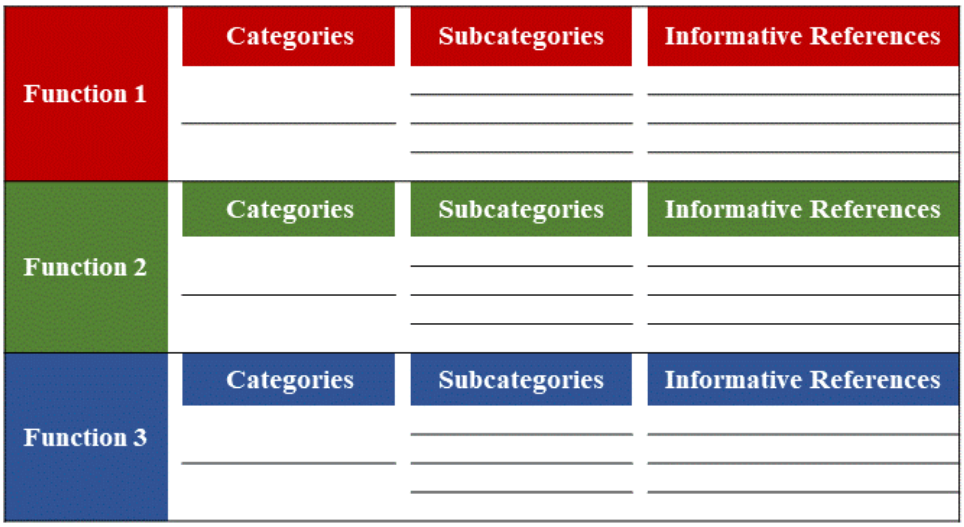


Fig. 2. Four elements of a Framework Core.

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

The Preliminary RDaF Framework Core is presented in Appendix E. The Core contains six Functions, which correspond to stages in the research data lifecycle, and Categories and Subcategories for each Function. The Functions are not intended to form a serial path or lead to a static desired end state. Rather, the Functions should be performed concurrently and continuously to create a dynamic operational culture that addresses the research data management needs. The six Functions (research data lifecycle stages) are defined below and depicted in Fig. 3.

- ***Envision*** – This Function encompasses the review of the overall strategies and drivers of an organization's research data program. The Envision Function is where choices and decisions are made that together chart a high-level course of action to achieve desired organizational goals. The Categories within this Function are Data Governance Structure, Community Engagement, Data Culture, Reward Structure, Workforce/Career Paths, Data Safety and Security, Strategy, and Data Risk Management.
- ***Plan*** – This Function encompasses the tactical management positioning in an organization for effective research data management throughout the research data lifecycle. The Categories within this Function are Chain of Control, Economics and Costs of Planning, Funding Planning, Data Objects, Hardware/Software Infrastructure, Data Management Planning, Scientific Data Standards, and Assessment and Controls.
- ***Generate/Acquire*** – This Function covers the 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. The Categories within this Function are Sources of Raw Data, Experimental Data Generation, Computational Data Generation, FAIR Principles for Data Generated In-House, External Sources of Data, and Community-Based Standards for Formats.
- ***Process/Analyze*** – This Function concerns the actions performed on generated or acquired research data to yield processed research data, typically using software, from which observations and conclusions can be made. This Function also concerns the research data stewardship functions performed by an organization. The Categories within this Function are Data Provenance, Data Architecture, Software Tools, Scientific Workflow Processes and Systems, Data Inventory, Data Modeling and Analytics, Data Representation/Models/Structures, Data Curation, and Metadata.
- ***Share/Use/Reuse*** – This Function outlines how raw and processed research data are disseminated, used, and reused within an organization and any constraints or encouragements to use/reuse. It also includes the dissemination, use, and reuse of raw and processed research data outside of an organization. The Categories within this Function are Legal and Licenses, Data Publishing, Data Citation, Internal and External Data Access, Levels of Protection, Applications and Analysis, and Data Architecture for Application and Use.
- ***Preserve/Discard*** – This Function delineates the end-of-use and end-of-life provisions for research data in an organization and includes records management, archiving, and safe disposal. The Categories within this Function are Criteria, Data Sustainability, Storage and Preservation of Data, Moving Data from One Service to Another across Organizations, and Retention and Disposition Schedules.

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

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

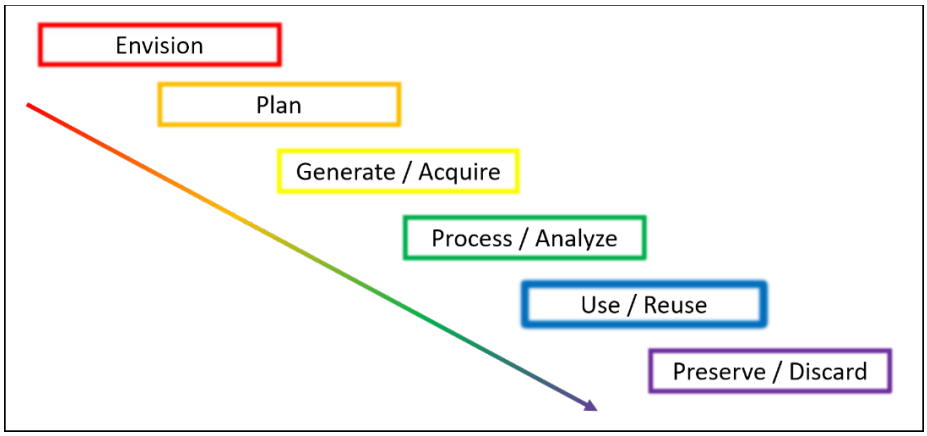


Fig. 3. The six Functions (research data lifecycle stages) of the Preliminary RDaF

3.3. Informative References

Informative References are typically standards, guidelines, and practices relevant to a specific Subcategory, but may also include laws, regulations, and other tools. These resources can support an organization's use of the RDaF to adopt better research data management practices. An initial list of Informative References relevant to the RDaF is given in Appendix F.

Mappings of Informative References to Subcategories provide implementation support, e.g., they help organizations determine which topics to prioritize to attain the desired state of research data management. A gap analysis of such mappings may also be used to identify where revised or additional standards, guidelines, and practices would help an organization to address emerging research data management needs.

3.4. Framework Profiles

Because the research data world is evolving so rapidly, new job functions such as data stewards and data scientists are emerging, and skilled people are in short supply [33, 34]. Guidelines and checklists to ensure that research data management considerations in the various roles are fully characterized and addressed are now a critical need. The concept of Framework Profiles, mentioned here for illustrative purposes, allows the RDaF to be tailored to different levels of stakeholders/users from a CEO to an individual researcher. To develop a Framework Profile, an organization can review all the Categories and Subcategories and determine which are relevant for an organizational unit and/or job function. Categories and Subcategories can be added as needed to fully adapt the RDaF to a specific stakeholder/user. Framework Profiles may be used to conduct self-assessments of research data management and communicate the results within an organization or between organizations. An example of Framework Profile development using a few Subcategories in the Envision Function/Data Governance Structure Category for various roles in an organization is provided in Table 3. This example clearly demonstrates that this subset of Data Governance topics is within the purview of a CDO.

Table 3. Example of Framework Profile development.

			Roles			
Function	**Category**	**Subcategory**	**CDO**	**Researcher**	**Librarian**	**Data Steward**
ENVISION Review of the overall strategies and drivers of an organization's research data program.	Data Governance Structure	Data vision and/or data policy	X			
		Legal and regulatory compliance	X			
		Data quality	X	X	X	X
		Data privacy	X	X	X	X
		Data management value proposition	X			X

3.5. Framework Implementation Tiers

Framework Implementation Tiers are not addressed in the Preliminary RDaF but will be included in the next version. Implementation Tiers will enable an organization to assess its current state of research data management and to develop a roadmap to attain its desired state of research data management. They can support an organization's decisions regarding research data management and help prioritize areas that would benefit from additional resources. For the RDaF, Implementation Tiers will be described in terms of data maturity, which has been defined as "the extent to which an organisation utilises the data they produce [35]" and "a measurement of the ability of an organization for continuous improvement in [data management] [36]." Maturity indicators are mechanisms to assess the extent of and success of research data management in an organization.

Maturity models define the fundamental processes of data management and specific capabilities and actions that constitute a path to improvements in data maturity. There are a large number of data management/governance maturity models, including DAMA-DMBOK2 [37], Data Management Capability Assessment Model (DCAM) [38], CCMI Institute Data Management Maturity Model [39], IBM Data Governance Council Maturity Model [40], Stanford Data Governance Maturity Model [41], Gartner Enterprise Information Management (EIM) Maturity Model [42] and EIM Framework [43], Social Security Administration Advanced Analytics Capability Maturity Model [44], and Federal Government Data Maturity Model [45].

4. Next Steps

The objective of the next phase in the development of the RDaF is to test the applicability and usefulness of the Framework Core in Appendix E. To accomplish this objective, two concurrent pilot studies—one in materials science and the other in research universities, including librarian and publisher roles—will be conducted. A timeline for the next phase is presented in Fig. 4. Because continuation of the RDaF effort is contingent on the availability of funding, the timeline begins with month zero. Prior to month zero, funding must have been secured and support staff identified.

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

Each pilot study will have three workshops, and each workshop will have "homework assignments" for the attendees and the broader pilot study community as described below:

1) "Kick-Off" Workshop: Attendees will be introduced to the Preliminary RDaF. Implementation of the RDaF will be discussed.
   Homework: *Community discusses how the Preliminary RDaF can help them with research data management and identifies their stakeholders and Informative References for the Preliminary Framework Core.*
2) "Working" Workshop: Attendees will report their findings and plan how to apply the RDaF.
   Homework*: Community tests the Preliminary RDaF and identifies refinements to it and Informative References for the Preliminary Framework Core.*
3) "Report" Workshop: Attendees will draft a report on the pilot study findings and discuss lessons learned.
   Homework: *Community completes their final report.*

   The RDaF Steering Committee will review the two pilot study final reports and revise the RDaF as needed. The next version of the RDaF will be released within six months of completion of the pilot study reports.

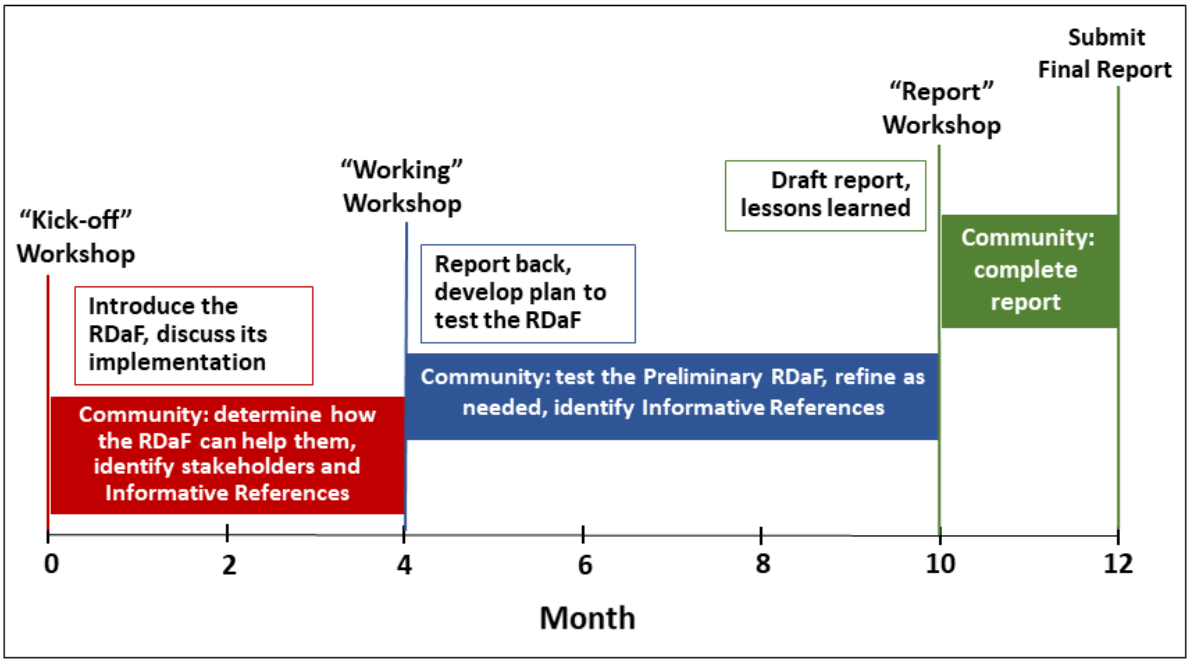


Fig. 4. Timeline for the two pilot studies.

Acknowledgments

NIST acknowledges and thanks all individuals who have contributed to this Preliminary RDaF. The authors acknowledge support provided by the NIST Material Measurement Laboratory for a contractor to assist in this work.

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

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

References

[1] Intangible property, 2 CFR § 200.315 (2014). Available at https://www.govinfo.gov/app/details/CFR-2014-title2-vol1/CFR-2014-title2-vol1-sec200-315

[2] Borek A, Parlikad AK, Webb J, Woodall P (2014) Data and Information Assets. *Total Information Risk Management: Maximizing the Value of Data and Information Assets,* eds Borek A, Parlikad AK, Webb J, Woodall P (New York, NY, Elsevier) Chapter 1, pp 3-22. https://doi.org/10.1016/C2012-0-00446-2

[3] Organization for Economic Co-operation and Development (2015) *Data-driven innovation for growth and well-being*. Available at https://web.archive.org/web/20201127113028/http://www.oecd.org/sti/ieconomy/data-driven-innovation.htm

[4] Holdren JP (2013 February 22) *Increasing Access to the Results of Federally Funded Scientific Research* [Memorandum] Executive Office of the President, Office of Science and Technology Policy. Available at https://web.archive.org/web/20201120104318/https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

[5] Royal Geographical Society (2015) *21st Century Challenges*. Available at https://web.archive.org/web/20210117160320/https://21stcenturychallenges.org/big-data-2/

[6] Currie ME, Paris BS, Donovan JM (2019) *What difference do data make? Data management and social change.* Online Information Review 43(6): 971-985. https://doi.org/10.1108/OIR-02-2018-0052

[7] Redman T (1998) The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM* 41(2):79-82. https://doi.org/10.1145/269012.269025

[8] Fruchterman J (2016) Using Data for Action and for Impact. Stanford Social Innovation Review. Available at https://web.archive.org/web/20201104180548/https://ssir.org/articles/entry/using_data_for_action_and_for_impact

[9] Rouse M, Stedman C, Vaughan J (October 2019) *What is data management and why is it important?* TechTarget. Available at https://web.archive.org/web/20201227230637/https://searchdatamanagement.techtarget.com/definition/data-management

[10] Wilkinson MD et al. (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(1):160018. https://doi.org/10.1038/sdata.2016.18

[11] European Union (2020) *European Open Science Cloud*. Available at https://web.archive.org/web/20201120103536/https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[12] Computer Network Information Center, Chinese Academy of Sciences (2021) *China Science and Technology Cloud.* Available at https://web.archive.org/web/20210117161142/https://www.cstcloud.net/news/23.jhtml

[13] Computer Network Information Center, Chinese Academy of Sciences (2021) *Global Open Science Cloud (GOSC)*. Available at https://web.archive.org/web/20210117161640/http://www.cstcloud.net/gosc.htm

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

[14] National Academies of Sciences, Engineering, and Medicine (2018) *Open Science by Design: Realizing a Vision for 21st Century Research* (National Academies Press, Washington, D.C.) Chapter 2, pp. https://doi.org/10.17226/25116

[15] National Academies of Sciences, Engineering, and Medicine (1997) Trends and Issues in Information Technology. *Bits of Power: Issues in Global Access to Scientific Data* (National Academies Press, Washington, D.C.) Chapter 2, pp. 24-46. https://doi.org/10.17226/5504

[16] Western Digital (2019 June 28) *How Supercomputers are Becoming a Scientist's Best Friend*. Available at https://web.archive.org/web/20210117161808/https://datamakespossible.westerndigital.com/supercomputers-becoming-scientists-best-friend/

[17] Network World (2020) *World's fastest supercomputers: Fugaku is still No. 1 at 3X the speed of No. 2*. Available at https://web.archive.org/web/20210117162204/https://www.networkworld.com/article/3597409/worlds-fastest-supercomputers-fugaku-is-still-no-1-at-3x-the-speed-of-no-2.html

[18] Wired (2019 January 1) *The World's Fastest Supercomputer Breaks an AI Record*. Available at https://web.archive.org/web/20210117162603/https://www.wired.com/story/worlds-fastest-supercomputer-breaks-ai-record/

[19] Saltmarch A (2020 April 16) *IBM's Summit—The Supercomputer Fighting Coronavirus*, Medical Expo E-Mag. Available at https://web.archive.org/web/20210117162921/http://emag.medicalexpo.com/summit-the-supercomputer-fighting-coronavirus/

[20] National Institute of Standards and Technology (2018) Framework for Improving Critical Infrastructure Cybersecurity, version 1.1 (U.S. Department of Commerce, Washington, D.C.), NIST Cybersecurity White Paper April 16, 2018. https://doi.org/10.6028/NIST.CSWP.04162018

[21] Collins S et al. (2018) *Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data* (European Commission, Brussels, EU). Available at https://web.archive.org/web/20201120103833/https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf

[22] Landi A et al (2020) The 'A' of FAIR – As Open as Possible, as Closed as Necessary. *Data Intelligence* 2(1-2): 47-55. https://doi.org/10.1162/dint_a_00027

[23] United States, Executive Office of the President Barack Obama (12 February 2013) Executive Order 13636: *Improving Critical Infrastructure Cybersecurity*. *Federal Register*, Vol. 78, No. 33, 19 Feb. 2013, pp. 11739-11744. Available at https://web.archive.org/web/20201120104102/https://www.govinfo.gov/content/pkg/FR-2013-02-19/pdf/2013-03915.pdf

[24] Burwell SM, VanRoekel S, Park T, Mancini DJ (2013 May 9) *Open Data Policy—Managing Information as an Asset* [Memorandum] Executive Office of the President, Office of Management and Budget. Available at https://web.archive.org/web/20201120104818/https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

United States, Executive Office of the President Barack Obama (2013 May 14) Executive Order 13642: *Making Open and Machine Readable the New Default for Government Information. Federal Register*, Vol. 78, No. 93, pp. 28111-28113. Available at https://web.archive.org/web/20201120105015/https://www.govinfo.gov/content/pkg/FR-2013-05-14/pdf/2013-11533.pdf

[25] Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. 115-435 §101 132 Stat. 5529 (2019). Available at https://web.archive.org/web/20201120105202/https://www.govinfo.gov/content/pkg/PLAW-115publ435/pdf/PLAW-115publ435.pdf

[26] OPEN Government Data Act, H.R. 1770, 115th Congress, 1st Session (2017). Available at https://web.archive.org/web/20201120105343/https://www.congress.gov/115/bills/hr1770/BILLS-115hr1770ih.pdf

[27] Office of Management and Budget, Office of Science and Technology Policy, Department of Commerce, Small Business Administration (2019) *Federal Data Strategy: Leveraging Data as a Strategic Asset*. Available at https://web.archive.org/web/20210117163443/https://strategy.data.gov/

[28] Commission on Evidence-Based Policymaking (2017) *The Promise of Evidence-Based Policymaking*. Available at https://web.archive.org/web/20201120105528/https://www.cep.gov/report/cep-final-report.pdf

[29] White House (2018) *The President's Management Agenda: Modernizing Government for the 21st Century*. Available at https://web.archive.org/web/20201120105737/https://www.whitehouse.gov/wp-content/uploads/2018/04/ThePresidentsManagementAgenda.pdf

[30] White House (2019) *The President's Management Agenda: Federal Data Strategy 2020 Action Plan*. Available at https://web.archive.org/web/20201120105948/https://strategy.data.gov/assets/docs/2020-federal-data-strategy-action-plan.pdf

[31] National Institute of Standards and Technology (2020) *NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management*, Version 1.0 (U.S. Department of Commerce, Washington, D.C.), NIST Cybersecurity White Paper January 16, 2020. https://doi.org/10.6028/NIST.CSWP.01162020

[32] National Institute of Standards and Technology (2019) *NIST Big Data Interoperability Framework*, version 3.0 (U.S. Department of Commerce, Washington, D.C.). Available at https://web.archive.org/web/20201120112757/https://bigdatawg.nist.gov/V3_output_docs.php

[33] DuBois J (2020 April 7) *The Data Scientist Shortage in 2020*. Quanthub. Available at https://web.archive.org/web/20210123140706/https://quanthub.com/data-scientist-shortage-2020/

[34] CompTIA (2020) *Trends in Data Management*. Available at https://web.archive.org/web/20210117163223/https://www.comptia.org/content/research/data-management-trends-survey

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

[35] Data Orchard (2020) *What is data maturity?* Available at https://web.archive.org/web/20201120113157/https://www.dataorchard.org.uk/what-is-data-maturity

[36] Steenbeek I (2020 February 10) *Data Management Maturity 101: What is a data management maturity assessment and why does a company need it?* Data Crossroad. Available at https://web.archive.org/web/20201209184621/https://datacrossroads.nl/2020/02/10/data-management-maturity-101-what-is-a-data-management-maturity-assessment-and-why-does-a-company-need-it/

[37] DAMA International (2017) *DAMA-DMBOK Data Management Body of Knowledge* (Technics Publishing, Basking Ridge, New Jersey), 2nd Ed.

[38] Enterprise Data Management Council (2018) *Data Management Capability Assessment Model (DCAM) Overview.* Available at *https://web.archive.org/web/20201120130740/https://cdn.ymaws.com/edmcouncil.org/resource/resmgr/featured_documents/EDMC_DCAM_Overview.pdf*

[39] CMMI Institute (2020) *Data Management Maturity (DMM) Model At-A-Glance*. Available at https://web.archive.org/web/20201120142150/https://cmmiinstitute.com/getattachment/cb35800b-720f-4afe-93bf-86ccefb1fb17/attachment.aspx

[40] Taylor K (2020 August 6) *Data Governance Maturity Models Explained*. Available at https://www.hitechnectar.com/blogs/data-governance-maturity-models-explained/

[41] Firican G (2018) *Stanford data governance maturity model*. Available at https://web.archive.org/web/20201217191145/https://www.lightsondata.com/data-governance-maturity-models-stanford/

[42] Newman D, Logan D (2008) *Gartner Introduces the EIM Maturity Model*. Available at https://web.archive.org/web/20201203142603/https://www.semanticscholar.org/paper/Gartner-Introduces-the-EIM-Maturity-Model-Newman-Logan/ca3b13f65a37d7b0a44287899710112e2c5afc4e

[43] White A (2014) *The Gartner Enterprise Information Management Framework*. Available at https://web.archive.org/web/20200326200148/https://blogs.gartner.com/andrew_white/files/2016/10/On_site_poster.pdf

[44] Social Security Administration Analytics Center of Excellence (2020) *Social Security Administration Advanced Analytics Capability Maturity Model*. Available at https://web.archive.org/web/20210117163110/https://www.ssa.gov/data/data_governance_board/ACE_A2CM2_for%20DGB.pptx.pdf

[45] Federal Government Data Cabinet (undated) *The Federal Government Data Maturity Model*. Available at https://web.archive.org/web/20210121170910/https://my.usgs.gov/confluence/download/attachments/624464994/Federal%20Government%20Data%20Maturity%20Model.pdf

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

Appendix A: Acronyms and Initialisms

AAU	Association of American Universities
AGU	American Geophysical Union
AI	Artificial Intelligence
ANDS	Australian National Data Service
APARD	Accelerating Public Access to Research Data
API	Application programming interface
APLU	Association of Public and Land-grant Universities
ARDC	Australian Research Data Commons
BRAIN	Brain Research through Advancing Innovative Neurotechnologies® [Initiative]
BRDI	Board on Research Data and Information
CDO	Chief Data Officer
CENDI	Commerce, Energy, NASA, Defense Information Managers Group
CEO	Chief Executive Officer
CMMI	Capability Maturity Model Integration
CNRI	Center for National Research Initiatives
CODATA	Committee on Data of the International Science Council
COI	Community of Interest
DAMA	Data Management Association International
DANS	Data Archiving and Networked Services
DCAM	Data Management Capability Assessment Model
DMBOK	Data Management Body of Knowledge
DMM	Data Management Maturity
DMP	Data Management Plan
DOC	Department of Commerce
DOE	Department of Energy
DOI	Digital Object Identifier
EIM	Enterprise Information Management
e-IRG	e-Infrastructure Reflection Group
ESFRI	European Strategy Forum on Research Infrastructures
ESIP	Earth Science Information Partners
EUDAT	European Data Infrastructure
FAIR	Findable, Accessible, Interoperable and Reusable
FORCE11	Future of Research Communications and e-Scholarship
GEIA	Government Electronics & Information Technology Association
GO FAIR	Global Open Findable, Accessible, Interoperable and Reusable
HPC	High-Performance Computing
HR	Human Resources
ICSTI	International Council for Scientific and Technical Information

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

IFLA	International Federation of Library Associations
ML	Machine Learning
NASEM	National Academies of Sciences, Engineering, and Medicine
NASA	National Aeronautics and Space Administration
NECTAR	Network for Effective Collaboration Technologies through Advanced Research
NFAIS	National Federation of Advanced Information Services (now merged with NISO)
NIH	National Institutes of Health
NISO	National Information Standards Organization
NIST	National Institute of Standards and Technology
NSF	National Science Foundation
NSTC	National Science and Technology Council
OECD	Organization for Economic Co-operation and Development
ORCID	Open Researcher and Contributor ID
OSTP	Office of Science and Technology Policy
RDA	Research Data Alliance
RDaF	Research Data Framework
SPARC	Scholarly Publishing and Academic Resources Coalition
SSP	Society for Scholarly Publishing
STM	International Association of Scientific, Technical and Medical Publishers
WDS	World Data System

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

Appendix B: Initial List of Stakeholders and Users

The following organizations and entities could potentially play a role in the development of the RDaF.

Private Funders

- Laura and John Arnold Foundation
- Alfred P. Sloan Foundation
- Bill & Melinda Gates Foundation
- Kavli
- Flat Iron
- Belmont Forum
- Helmsley Charitable Trust
- Wellcome Trust

Data Centers

- World Data System (WDS) and its members (particularly U.S. member centers)

Repositories and Service Providers

- re3data (Registry of Research Data Repositories)
- DataONE (Data Observation Network for Earth)
- Figshare
- Dryad
- DataCite
- ORCID (Open Researcher and Contributor ID)

Library and Not-for-Profit Organizations

- California Digital Library - DMPTool
- National Information Standards Organization (NISO)
- Association for Research Libraries
- Scholarly Publishing and Academic Resources Coalition (SPARC)
- Center for Open Science
- Center for National Research Initiatives (CNRI)
- International Federation of Library Associations (IFLA)

University Organizations

- Association of American Medical Colleges
- Association of American Universities (AAU)
- Association of Public and Land-grant Universities (APLU)

Publishing Community

- Elsevier
- Springer Nature
- Society for Scholarly Publishing (SSP)
- Coalition for Publishing Data in the Earth and Space Sciences
- International Association of Scientific, Technical and Medical Publishers (STM)

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

Data Organizations

- Committee on Data of the International Science Council (CODATA)
    - NASEM: U.S. National Committee for CODATA, associated with the NASEM Board on Research Data and Information (BRDI)
- Research Data Alliance (RDA)
    - Special focus on RDA-U.S.
- Earth Science Information Partners (ESIP)
- Esri, formerly Environmental Systems Research Institute
- International Council for Scientific and Technical Information (ICSTI)
- National Federation of Advanced Information Services (NFAIS), now merged with NISO
- Future of Research Communications and e-Scholarship (FORCE11)
- Commerce, Energy, NASA, Defense Information Managers Group (CENDI)
- Global Open Findable, Accessible, Interoperable and Reusable (GO FAIR)

Disciplinary/Topical Initiatives

- Materials Genome Initiative
- BRAIN Initiative
- Integrated Global Greenhouse Gas Information System
- Biodiversity Global Information Facility
- American Geophysical Union (AGU)
- Accelerating Public Access to Research Data (APARD)

Federal Agencies and Programs

- National Institute of Standards and Technology (NIST)
- Department of Energy (DOE)
- National Optical Astronomy Observatory
- National Aeronautics and Space Administration (NASA)
- National Institutes of Health (NIH)
- National Library of Medicine

Policy/Studies Organizations

- National Science and Technology Council (NSTC)
    - Subcommittee on Open Science, formerly the Interagency Working Group on Open Science
- National Academies of Sciences, Engineering, and Medicine (NASEM)
    - Board on Research Data and Information (BRDI)

International Agencies and Programs

- Organization for Economic Co-operation and Development (OECD)
- European Data Infrastructure (EUDAT)
- European Open Science Cloud
- International Science Council
- China Science and Technology Cloud
- International Bureau of Weights and Measures (BIPM)
- e-IRG – e-Infrastructure Reflection Group

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

Foreign Governments and National Organizations

- Australian Research Data Commons (ARDS), a merger between ANDS, National eResearch and Collaboration Tools and Resources, and Research Data Services, Australia
- Commonwealth Scientific and Industrial Research Organisation (Australia)
- CANAIRE, formerly the Canadian Network for the Advancement of Research, Industry, and Education
- Data Archiving and Networked Services (DANS, the Netherlands)
- Academy of Science of South Africa
- International Development Research Center (Canada)
- Economic Commission for Latin America and the Caribbean
- São Paulo Research Foundation (Brazil)
- European Strategy Forum on Research Infrastructures (ESFRI)

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

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

Appendix C: RDaF Stakeholder Scoping Workshop Agenda



AGENDA

DAY 1: Dec. 5 *"Why an RDaF?"*

Time	Topic	Comments
8:30-9:00	CONTINENTAL BREAKFAST/NETWORKING	
9:00-9:30	Welcome	**Walter Copan** (NIST Director) **Jim St. Pierre** (ITL Deputy Director; Setting the Framework context at NIST)
9:30-9:45	***Setting the stage***	**Bob Hanisch** *(Director, Office of Data and Informatics, MML, NIST)* What are we trying to achieve at this workshop? What is different about the RDaF from the other NIST Frameworks? Who do we see as stakeholders and beneficiaries? What do we all want to see in Framework and why? How shall we proceed?
9:45-10:30	***Understanding the Ecosystem:*** - Stakeholder perspectives	This will be flash talks (5-6 minutes each) from different categories of stakeholders to tee up break-outs. Present who players are; how to engage them; why are they there; what's their role; sustainability/sustainability model. Six minutes each; no slides.
10:30-10:45	BREAK	
10:45-11:45	Break out: ***Users and beneficiaries?***	What would be the value of an RDaF Why would you would want a Framework. How could it be of value? (Ideas like RDaF's are expensive. Do we really know that? Realistic estimate of costs – could we explore that? Costs of not having it?) Explore users and beneficiaries; these maybe not the exactly the same.
11:45-12:30	Group Presentations and Discussion	RESULTS: 1) Who are users & beneficiaries; 2) Use cases; 3) What stakeholder groups aren't here that should be included?
12:30-1:30	LUNCH with table discussions on morning topics	
1:30-2:15	***Understanding Life Cycle:*** - Stakeholder Perspectives	This will be 5 flash talks (6 minutes) from stakeholders who represent different functions in the data life cycle to tee up break-outs. Objective is to understand how the Framework might address layers or functions in the life cycle.
2:15-3:15	Break-out: ***Applications***	Taking a life cycle perspective What else do you want from an RDaF and identify additional use cases.
3:15-3:30	BREAK	
3:30-4:15	Group Presentations and Discussion	RESULTS: 1) What is wanted from a RDaF? 2) Additional User Cases for a Framework
4:15-5:00	***RDaF Contents***	Brief Presentations on what might be in an RDaF to tee up the morning break-outs.
6:00-	Informal Networking/Group Dinners	Informal gathering at the open Bar in the Courtyard by Marriott Rockville 2500 Research Blvd, Rockville, MD 20850. Groups can self-organize for dinner at local restaurants.

DAY 2: Dec. 6 -- *"What should be in it and how to develop it?"*

Time	Topic	Comments
8:30	**GET YOUR COFFEE**	
8:40-9:00	Recap of Day 1	We've gained an understanding of many perspectives and how it can be used and some use cases
9:00-10:00	Break out: ***Content, Components, and Elements*** -What should be in an RDaF?	Pick 2-3 use cases to help guide thinking, but don't be restricted to them. Based on what we've heard and the use cases what should be in an RDaF? What components need to be included to come out supporting use cases and to deliver benefits?
10:00-10:45	Report out & discussion – Developing RDaF content.	Begin to develop a conceptual outline of an RDaF.
10:45-11:00	**BREAK**	
11:00-11:45	Break-out: ***Process.*** -How should we go about it?	How should we proceed from here to complete the scoping study? (Provide the proposed steps?) Assuming a positive outcome of the scoping, what should be the next steps? What is proposed: 1) full scoping 2)Pilot in two disciplines 3)Full development? Who should be involved?
11:45-12:30	Group Presentations and Discussion	RESULTS: Each group should present 1) a path to take the workshop results and complete the scoping. Once the initial scoping is done, 2) what should be the steps to a final Framework and what would it take to get there.
12:30-1:00	Summary and wrap up	What have we heard and where will we go from here? What is the interest in continued involvement of participants?
1:00-3:00	Steering Group Meeting	The Steering Group will convene for a working lunch.

FLASH TALK PRESENTERS
Stakeholder perspectives

Time	Topic	Speakers
9:45-10:30	***Understanding the Ecosystem***: -	Government: **Laura Biven**, Department of Energy Academia: **Sayeed Choudhury**, Johns Hopkins University Industry: **Vivien Bonazzi**, Deloitte International: **Jean-François Abramatik**, European Open Science Cloud Community: **Hilary Hanahoe**, Research Data Alliance Chief Data Officer: **Ed Kearns**, Department of Commerce
1:30-2:15	***Understanding Life Cycle:***	Funder: **Beth Plale**, National Science Foundation Researcher: **Barend Mons**, Leiden University Medical Center & CODATA Publisher: **Shelley Stall**, American Geophysical Union Infrastructure: **Mark Leggott**, Research Data Canada Library/Archive: **Leah McEvan**, Cornell University
4:15-5:00	***RDaF Contents***	**Anita deWaard**, Elsevier **Merce Crosas**, Harvard **Susan Gregurick**, National Institutes of Health

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

Appendix D: RDaF Stakeholder Scoping Workshop Attendees

Last Name	First Name	Organization
Abramatic	Jean-François	National Institute for Research in Computer Science and Automation, France
Agarwal	Deborah	Lawrence Berkeley National Laboratory, DOE
Allard	Suzanne	University of Tennessee, Knoxville
Ananthakrishnan	Rachana	University of Chicago
Ang	James	Pacific Northwest National Laboratory, DOE
Biven[a]	Laura	Office of Science, DOE
Bonazzi	Vivien	Deloitte
Bruce	Elizabeth	Microsoft
Carroll	Bonnie	Information International Associates and CODATA
Choudhury	Golam	Johns Hopkins University
Cragin	Melissa	University of California, San Diego
Crosas	Mercé	Institute for Quantitative Social Science, Harvard University
Dahlitz	Karen	Australia (no affiliation provided)
de Waard	Anita	Elsevier
Dreisigmeyer	David	U.S. Census Bureau, DOC
Erdmann	Christopher	Renaissance Computing Institute, University of North Carolina
Fagnan	Kirsten	Lawrence Berkeley National Laboratory, DOE
Federer	Lisa	National Library of Medicine, NIH
Govoni	Marco	Argonne National Laboratory, DOE
Gregurick	Susan	Office of Data Science Strategy, NIH
Hanahoe	Hilary	Research Data Alliance, Italy
Hanisch	Robert	Material Measurement Laboratory, NIST
Hanson	Brooks	American Geophysical Union
Honaker	James	Center for Research on Computation and Society, Harvard University
Hudson-Vitale	Cynthia	Association of Research Libraries
Johnston	Lisa	University of Minnesota
Kahn	Scott	LunaDNA
Kaiser	Debra	Material Measurement Laboratory, NIST
Kearns	Edward	DOC
Kitney	Stuart	National Physical Laboratory, United Kingdom
Leggott	Mark	Research Data Canada
Lucas	Matthew	Social Sciences and Humanities Research Council of Canada
McEwen	Leah	Cornell University
Medina-Smith	Andrea	Information Services Office, NIST

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

Last Name	First Name	Organization
Mons	Barend	CODATA, GO-FAIR, and Leiden University, Netherlands
Musen	Mark	Stanford University
Nichols	Lisa	Office of Science and Technology Policy
Plale[b]	Beth	National Science Foundation
Pollard	Tom	Massachusetts Institute of Technology/PhysioNet Repository
Pouchard	Line	Brookhaven National Laboratory, DOE
Ricci	James	Advanced Scientific Computing Research, DOE
Robinson	Carly	Office of Scientific and Technical Information, DOE
Schlenoff	Craig	Program Coordination Office, NIST
Sellars	Scott	Department of State
Shyam Sunder	Sivaraj	Laboratory Programs, NIST
Stall	Shelley	American Geophysical Union
Strawn	George	BRDI, NASEM
Uhlir	Paul	Self-employed
Vanderwall	Dana	Bristol-Myers Squibb and Allotrope Foundation
Woo	Kara	Sage Bionetworks

[a]Now at Office of Data Science Strategy, NIH.
[b]Now at Indiana University.

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

Appendix E: Preliminary RDaF Framework Core

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.

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
ENVISION Review of the overall strategies and drivers of an organization's research data program.	Data **Governance*** Structure	• Identification of Goals and Roles • Data vision and/or **data policy** • Data management value proposition • Data management organization • Value of data (quantitative or qualitative) • Legal and regulatory **compliance** • **Data quality** (including Trust and Certification) • **Data privacy** • Data ethics
	Community Engagement	• **Stakeholder community(ies)** • Communication with stakeholder community(ies) • Interactions with other organizations • **Cross-community engagement** (across domains and sectors) • Inclusivity in interactions
	Data Culture*	• **FAIR data principles** • **Value of data** • Roles and responsibilities
	Reward Structure	• For data management • Value of **data workers** • **Incentives and institutional credit for data sharing** and reuse • Disincentives for data sharing • Human Resources (HR) involvement

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

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
ENVISION (continued)	**Workforce/Career Paths**	• 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 **Safety and Security**	• Safety and security assurance • Data inventory
	Strategy	• Organizational data management
	Data **Risk** Management*	• Risk assessment • Risk mitigation and management
PLAN The tactical management positioning in an organization for effective research data management throughout the research data lifecycle*.	Chain of Control	• Documentation • Communication within organization
	Economics and Costs of Planning	• Decision-making tools for data, including cost-benefit analysis • Cost breakdown, i.e., calculation of costs by data lifecycle* stage
	Funding Planning	• Models for **provisioning resources**, i.e., direct, overhead, or mixed
	Data Objects	• Quantitative and qualitative data • Software, models • Instruments • Data publications*, journal publications • Presentations • Other
	Hardware/Software Infrastructure	• Interoperability • Persistent instrument identifiers
	Data Management Planning	• **Data management plans** (DMPs) • Lifecycle considerations: living documents or static proposals?
	Scientific Data Standards	• Sources of standards • General, domain-specific
	Assessment and Controls	• Goals/definition of success • **Metrics** or metrics structure, tracking use and impact measures

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

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
GENERATE/ ACQUIRE The generation of raw research data and/or the acquisition* of research data by an organization.	Sources of Raw Data*	• Generated In-house experimentally or computationally • Collected from external sources
	Experimental Data Generation	• Specification and recording of instruments and associated metadata • Description and recording of measurement protocols • Methods for data and metadata capture and recording
	Computational Data Generation	• Commercial and/or custom software • Methods for computational variables (metadata) capture and recording
	FAIR Principles for Data Generated In-House	• Data born FAIR • Data made FAIR
	External Sources of Data	• Data acquired FAIR • **Identification, collection, and recording** • **Metadata harvesting**
	Community-Based Standards for Formats	• Standards development organizations/sources • General, domain-specific
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.	Data **Provenance**	• Original authoritative copy • Version identification • Provenance of data derived from other data • Provenance of scientific records across all the individual outputs • **Timestamping**
	Data Architecture	• Design • Security • Configuration management • Hosting and storage • Use of cloud
	Software Tools	• Data lifecycle* • Management and analysis • Commercial and/or custom tools • System resilience and adaptability • Maintenance

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

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

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
PROCESS/ ANALYZE (continued)	Scientific **Workflow** Processes and Systems	• Workflow tools • Laboratory notebooks, i.e., electronic, paper
	Data Inventory	• Formats and standards • Catalogs • Interoperability (across instrument manufacturer file formats)
	Data Modeling and Analytics	• Processes • Tools
	Data Representation/ Models/Structures	• Dynamic data • General, domain-specific
	Data Curation	• Policies and processes • Manpower
	Metadata	• Types of metadata • Responsible parties • Specification of metadata standards • **Linked data** structure • Persistent identification (DOI)
SHARE/USE/ REUSE How research data are disseminated, used, and reused within and outside an organization.	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*
	Data Publishing*	• Repositories • Referencing data/digital objects from journal articles • Supplementary material • Data linking
	Data Citation*	• Citation metrics • Citation impact
	Internal and External Data Access	• Access internally, e.g., the data generator • Access externally • **Programmatic access**, aka Smart API • Data access vs. data visiting
SHARE/USE/	Levels of Protection	• Unclassified but sensitive information, e.g., de-identification, enclaves

FUNCTION (Data Lifecycle* Stage)	CATEGORY	SUBCATEGORY
REUSE (continued)	Levels of Protection (continued)	• Security classification • Protecting limited data/secure platforms/enclaves • Data anonymization*
	Applications and Analysis	• **Technologies** for use and analytics, e.g., AI, ML
	Data Architectures for **Application and Use**	• Extensibility across communities, including machine-based interactions • Capturing insights from ML and use of these to improve datasets for future AI applications • Capturing data performance characteristics • Location of data (e.g., relative to instruments, in the cloud, transient copies)
PRESERVE/ DISCARD The end-of-use and end-of-life provisions for research data in an organization, including records management, archiving, and safe disposal.	Criteria	• Use and impact
	Data Sustainability	• Data longevity and support • Orphan datasets
	Storage and Preservation of Data	• Media to store and preserve data • Data back-up • Data repositories
	Moving Data from One Service to Another across Organizations	• Roles and responsibilities • Moving data from one agency to another, e.g., from a funded research agency to an agency with a permanent repository • Registration of repositories: roles and responsibilities • Disciplinary archives

	Retention and Disposition Schedules	• Data **archiving**, i.e., **what is kept and not kept**
		  • Decision processes
		• **End-of-life** issues
		  • Example: Responsible party for keeping raw data* feeds
		  • Example: Store (or not) raw data*, given the large amount of storage needed
		• Deaccessioning/End-of-life
		• Recognition of removed data (gravestone)

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

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

Appendix F: Initial List of Informative References

NIST Frameworks

- *Framework for Improving Critical Infrastructure Cybersecurity*, version 1.1 (2018). https://doi.org/10.6028/NIST.CSWP.04162018
- *NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management*, Version 1.0 (2020). https://doi.org/10.6028/NIST.CSWP.01162020
- *NIST Big Data Interoperability Framework*, V3.0 Final Version (2019). Available at https://web.archive.org/web/20201120112757/https://bigdatawg.nist.gov/V3_output_docs.php

Some Other Frameworks

- ANDS (Australian National Data Service)
  - Creating a Data Management Framework (2018). Available at https://web.archive.org/web/20200712212141/https://www.ands.org.au/__data/assets/pdf_file/0005/737276/Creatinga-data-management-framework.pdf
- DAMA (Data Management Association International)
  - Data Management Body of Knowledge Book (DMBOK2), 2nd ed. (2017). Available at https://web.archive.org/web/20201202151547/https://www.dama.org/content/what-data-management
  - DAMA-DMBOK2 Framework (2014). Available at https://web.archive.org/web/20201202165515/https://www.datasqlvisionary.com/wp-content/uploads/2018/06/DMBOK-Framework.pdf
- NISO (National Information Standards Organization)
  - Research Data Management (2015). Available at https://web.archive.org/web/20201202170023/https://groups.niso.org/apps/group_public/download.php/15375/PrimerRDM-2015-0727.pdf
- CMMI Institute
  - Data Management Maturity (DMM) Model (2019). Available at https://web.archive.org/web/20201120142150/https://cmmiinstitute.com/getattachment/cb35800b-720f-4afe-93bf-86ccefb1fb17/attachment.aspx

Guidelines and Initiatives

- AAU-APLU Public Access Working Group Report and Recommendations (2017). Available at https://web.archive.org/web/20200829090826/https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Intellectual-Property/Public-Open-Access/AAU-APLU-Public-Access-Working-Group-Report.pdf
- Redd K, Steen K, Nusser S, Smith T, Walters T, Chasen J, Luther J, and Reecy J (2020) *APLU AAU Accelerating Public Access to Research Data Workshop Report*. https://doi.org/10.31219/osf.io/63mxh

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

- AAU-APLU Accelerating Public Access to Research Data Summit. (2020 February 20-21). Available at https://web.archive.org/web/20201203130842/https://www.aau.edu/national-summit-accelerating-public-access-research-data

Articles, Reports, and Presentations

- Steenbeek I (2018) *Data Management maturity models: a comparative analysis.* Data Crossroads. Available at https://web.archive.org/web/20201203131916/https://datacrossroads.nl/2018/12/16/data-management-maturity-models-a-comparative-analysis/
- Proença D, Borbinha J (2018) Maturity models for data and information management. *Digital Libraries for Open Knowledge: 2nd International Conference on Theory and Practice of Digital Libraries*, eds Méndez E, Crestani F, Ribeiro C, David G, Lopes JC (Springer International Publishing, New York, New York) Vol. 11057, pp. 81-93. https://doi.org/10.1007/978-3-030-00066-0_7
- Mecca M (2014) *CMMI Institute Data Management Maturity Model Introduction*. Available at https://web.archive.org/web/20201203180752/https://cdn.ymaws.com/www.globalaea.org/resource/collection/68814379-BF7E-41C8-B152-18A617F9C0AA/Data_Management_Maturity_Model_Introduction_-_Dec_12_2014.pdf
- RDA FAIR Data Maturity Model Working Group (2020) FAIR Data Maturity Model: specification and guidelines. *Research Data Alliance*. https://doi.org/10.15497/RDA00050
- Enterprise Data Management Council (2018) *Data Management Capability Assessment Model (DCAM) Overview*. Available at https://web.archive.org/web/20201120130740/https://cdn.ymaws.com/edmcouncil.org/resource/resmgr/featured_documents/EDMC_DCAM_Overview.pdf
- Enterprise Data Management Council (2014) *Data Management Capability Assessment Model (DCAM)*, *Working Draft Version 7*. Available at https://web.archive.org/web/20201204101045/https://dgpo.org/wp-content/uploads/2016/06/EDMC_DCAM_-_WORKING_DRAFT_VERSION_0.7.pdf
- SAE International, EIDM Enterprise Information and Data Management Committee (2012) *Data Management Aerospace Standard* (GEIA859A). https://doi.org/10.4271/GEIA859A

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

Appendix G: Glossary of Terms used in Appendix E

Data acquisition "The process of acquiring data from some source. For example, data may be acquired by download from a repository, transfer from a data logger, data capture, etc. [1]."

Data anonymization Data anonymization is a type of information sanitization intended to protect privacy. It is a "process by which personal data is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party [2]." "Data anonymization may enable the transfer of information across a boundary, such as between two departments within an agency or between two agencies, while reducing the risk of unintended disclosure, and in certain environments in a manner that enables evaluation and analytics post-anonymization [3]."

Data citation "Data citation is the provision of accurate, consistent, and standardized referencing for datasets just as bibliographic citations are provided for other published sources like research articles or monographs. Typically, the well-established Digital Object Identifier (DOI) approach is used with DOIs taking users to a website that contains the metadata on the dataset and the dataset itself [4]."

Data culture "Data culture is the principle established in the process of social practice in both public and private sectors which requires all staffs and decision-makers to focus on the information conveyed by the existing data and make decisions and changes according to these results instead of leading the development of the company based on experience in the particular field [5]."

Data governance "The policies, procedures, and processes to manage and monitor the organization's regulatory, legal, risk, environmental, and organizational requirements are understood and inform the management of [data] risk [6]."

Data lifecycle "Refers to all the stages in the existence of digital information from creation to destruction. A lifecycle view is used to enable active management of the data objects and resource over time, thus maintaining accessibility and usability [7]."

Data publication "The release of research data, associated metadata, accompanying documentation, and software code (in cases where the raw data have been processed or manipulated) for re-use and analysis in such a manner that they can be discovered on the Web and referred to uniquely and persistently. Data publishing occurs via dedicated data repositories and/or (data) journals which ensure that the published research objects are well documented, curated, archived for the long term, interoperable, citable, quality-assured, and discoverable – all

aspects of data publishing that are important for future reuse of data by third-party end-users [8]."

Data stewardship "The most common label to describe accountability and responsibility for data and processes that ensure effective control and use of data assets. Stewardship can be formalized through job titles and descriptions, or it can be a less formal function driven by people trying to help an organization get value from its data [9]."

Raw data "Data that have not been processed for meaningful use. Although raw data have the potential to become "information," they require selective extraction, organization, and sometimes analysis and formatting for presentation. As a result of processing, raw data sometimes end up in a database, which enables the data to become accessible for further processing and analysis in several different ways [10]."

Risk management "Risk management refers to the practice of identifying potential risks in advance, analyzing them, and taking precautionary steps to reduce/curb the risk [11]." "Data carries tremendous value for organizations while creating new challenges around transparency, accuracy, security, privacy, social expectations, and legal requirements [12]."

References for Appendix G

[1] CASRAI Research Data Management Glossary (2020) *Data acquisition*. Available at https://web.archive.org/web/20201205144242/https://casrai.org/term/data-acquisition/

[2] International Organization for Standardization (2017) *Health informatics – Pseudonymization* (ISO Standard No. 25237:2017). Available at https://web.archive.org/web/20201206165253/https://www.iso.org/standard/63553.html

[3] Wikipedia (2020) *Data anonymization*. Available at https://web.archive.org/web/20201206175040/https://en.wikipedia.org/wiki/Data_anonymization

[4] Wikipedia (2020) *Data citation*. Available at https://web.archive.org/web/20201206180554/https://en.wikipedia.org/wiki/Data_citation

[5] Wikipedia (2020) *Data culture*. Available at https://web.archive.org/web/20200812184552/https://en.wikipedia.org/wiki/Data_culture

[6] National Institute of Standards and Technology (2018) Framework for Improving Critical Infrastructure Cybersecurity, version 1.1 (U.S. Department of Commerce, Washington, D.C.), NIST Cybersecurity White Paper April 16, 2018. https://doi.org/10.6028/NIST.CSWP.04162018. On pp.25-26, in Identify Function, Governance Category, the word "cybersecurity" was replaced with the word "data."

[7] CASRAI Research Data Management Glossary (2020) *Data lifecycle*. Available at https://web.archive.org/web/20201205145404/https://casrai.org/term/data-lifecycle/

[8] CASRAI Research Data Management Glossary (2020) *Data publication*. Available at https://web.archive.org/web/20201205150232/https://casrai.org/term/data-publication/

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

[9] DAMA International (2017) *DAMA-DMBOK Data Management Body of Knowledge* (Technics Publishing, Basking Ridge, New Jersey), 2nd Ed, pp. 75-76.
[10] CASRAI Research Data Management Glossary (2020) *Raw data*. Available at https://web.archive.org/save/https://casrai-test.evision.ca/glossary-term/raw-data/
[11] Definition of Risk Management (2020) *The Economic Times.* Available at https://web.archive.org/web/20201209184047/https://economictimes.indiatimes.com/definition/risk-management
[12] Albinson N, Thomas C, Rohrig M, Chu Y (2019) Future of risk in the digital era. *Deliotte*. Available at https://web.archive.org/web/20201209183838/https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-rfa-future-of-risk-in-the-digital-era-report.pdf

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