RDA-OfR Mapping the landscape of digital research tools WG

1. Charter

The digital research data infrastructure landscape comprises a myriad of [tools](#)[1] for managing and sharing research data during various stages of the research data lifecycle. Such research tools vary widely depending on data type, user requirement, provider, and subject area[2]. In the context of this WG, research 'tools enable researchers to perform one or more operations, typically on data, and often with data as the output. Tools are usually intended for use by humans. In this context we are explicitly excluding physical instruments.'[3]

The diversity and variety of research tools can prove overwhelming and challenging for stakeholders working within the digital research data ecosystem to understand, navigate, and select the most appropriate tool to meet their needs and objectives. The categorisation of research tools, based on their features, functionalities and how they interoperate, remains unclear. In many cases, research tools are not interoperable, often leading to siloed working within organisations and disciplines, thereby limiting the scope of research and the ability to share and reuse data.

This RDA Working Group (WG), supported by Oracle for Research (OfR), aims to address these challenges by: (i) categorising different types of research tools; and, (ii) mapping different types of research tools to the research data lifecycle based on their features and functionalities.

The WG will produce a categorisation schema (a conceptual framework) of research tool types that includes terminology, definitions and associated metadata describing features and functionalities of different tool types. The categorisation schema will be stored in an [autonomous database](#) provided by [Oracle Cloud Infrastructure](#). The WG will undertake the following programme of work to achieve its deliverables:

The creation of a research data lifecycle model and crosswalk to existing models (Deliverable 1)

The WG will examine and identify the different stages of the research data lifecycle. Since numerous different models of the research data lifecycle exist that have been conceptualised for specific research paradigms and audiences, the WG will conduct a landscape review to

---

[1] CODATA (2022) Research Data Management Terminology. Available at: https://codata.org/initiatives/data-science-and-stewardship/rdm-terminology-wg/rdm-terminology/ (Accessed: 30 May 2023).

[2] DARE UK Consortium. (2021). UK Data Research Infrastructure Landscape. Available at: https://doi.org/10.5281/zenodo.5584696. (Accessed: 30 May 2023).

[3] Jones, S., Leggott, M., Lopez Albacete, J., Madalli, D., Pascu, C., Payne, K., Schouppe, M., & Treloar, A. (2023). GORC IG: Typology and Definitions (Version 0.9.1). Research Data Alliance. https://doi.org/10.15497/RDA00087

research and consult existing models (see Section 3) and identify common stages of the research data lifecycle for use as the framework to guide the research tool categorisation. Each research data lifecycle stage will be supported by terminology and definitions. The WG will create a crosswalk to demonstrate connections between the chosen model and existing models.

The identification, categorisation, and mapping of different types of research tools: A categorisation schema (Deliverable 2)

The WG will research and consult existing work in the area to identify, categorise, and map different types of research tools. Such tools may include, but are not limited to: open science frameworks, data management planning tools, electronic laboratory notebooks (ELNs), laboratory information management systems (LIMS), virtual research environments (VREs), databases, repositories, and archives. Types of research tools will be described, categorised, and mapped to the research data lifecycle framework based on their utility, and assessed based on their interoperability.

The aim of this deliverable is to highlight the potential for and current limitations of streamlined flow of research data and metadata throughout the research data lifecycle based on how different types of research tools interoperate. This will be highly valuable in the context of the development of the national and international open research commons.

This work will contribute to and build on the work of the RDA's Global Open Research Commons IG and GORC International Model WG. Task Group 5 of the GORC International WG has undertaken an extensive literature review and released a Commons Attributes Model (Version 0.5) that identifies a suite of services and tools that will inform the work of this WG. Efforts to describe the features, functionality, and interoperability of different types of research tools will complement the development of the 'Commons Integration Roadmap' (GORC WG Deliverable) by providing key information about different types of research tools, and highlighting areas for the improvement of their interoperability and user experience.

The creation of a preliminary structural framework for an online open access 'map of the digital research tool landscape' (Deliverable 3)

The WG will undertake the necessary foundational work required to create an autonomous relational database that is hosted by the RDA Foundation (as a legal entity on behalf of the RDA), owned by the community, and powered by Oracle for Software. This arrangement has been discussed and agreed by RDA and Oracle for Research.

The open access database, navigable by research data lifecycle stage, will: (i) contain searchable information (e.g., features, functionalities, interoperability) about different types of research tools; and: (ii) allow for ongoing community curation and further development. The WG will provide recommendations for the long-term maintenance, sustainability, and adoption of the database to ensure that it remains current, relevant, and useful for the

research data community. Such recommendations will also propose methodologies for future community-curation (detailing *who* can contribute and *how*), management, and governance of the database.

The ultimate goal is to provide the research data community with a dynamic resource that remains up to date with newly emerging types of research tools and evolves with the ever-changing digital research data infrastructure landscape. This may include significant data and software-related developments, e.g., Artificial Intelligence (AI).

2. Value Proposition

To our knowledge, this RDA WG is the first initiative of its kind to categorise different types of research tools with a primary focus on their utility and interoperability within the research data lifecycle. Providing the global research data community with a high-level map of the digital research data tool landscape that can be navigated according to specific data management and sharing related tasks represents a novel approach to characterising the research data ecosystem. The outputs and recommendations produced by this WG aim to provide value and impact for the following adopters:

Adopter	Value/Impact
Researchers (e.g., data creators and users)	To understand, navigate, and select suitable research tools for managing and sharing data by providing information about their functionalities, relevance, and applicability to the various stages of the research data lifecycle.
Data support professionals (e.g., data managers)	To gain improved understanding of the digital research data infrastructure landscape, and become better equipped with essential knowledge of different types of research tools to provide relevant support, training, and education.
Open Science/Research/Data Commons professionals	To understand the features, functionalities, and interoperability of different types of research tools that can be used within diverse marketplaces or 'commons' for data and services.
Tool developers/ providers	To: (i) understand the different research tools operating within the digital research data landscape; and, (ii) improve tool features, functionalities, harmonisation, and interoperability to enhance data management and sharing practice.
Research performing organisations	To make informed recommendations at the organisational policy level to staff regarding appropriate types of research tools for the management and sharing of research data.
Publishers	To make informed recommendations to authors and journal editors regarding appropriate types of research tools for the management, publication, and sharing of data associated with journal manuscripts.
Funders	To make informed recommendations to researchers and project managers based on data management plans for funded research.

3. Engagement with existing work in the area

This working group contributes to and builds on a number of preceding and existing initiatives (e.g., frameworks, registries, and directories) that signpost or aggregate tools within the digital research data infrastructure landscape. However, most initiatives to date focus on specific: (i) regions/nations; (ii) disciplines; or (iii) research tools (primarily databases and repositories), providing a high level of granularity.

This WG aims to build on and contribute to existing work in the area by creating a high-level map (a 'birds-eye view') of the digital research tool landscape. To achieve its proposed programme of work (outlined in Section 1), the WG will engage with the following organisations, projects, and initiatives:

Please note this is not an exhaustive list and the WG may find more examples of relevant existing work to include during the initial research and consultation phase.

For the creation of a research data lifecycle framework and crosswalk to existing models:

- NIST Research Data Framework (RDaF), specifically Version 1.5[4] - Provides a map of the research data space that uses a lifecycle approach with six high-level lifecycle stages, topics, and subtopics to organise key information concerning RDM and research data dissemination.
- DCC Curation Lifecycle Model - A data-centric model that defines research data management workflows and associated roles and responsibilities within an organisation.
- ARDC Research Data Management Framework for Institutions - Australian national framework that features 19 essential elements for research data management.

For the identification, categorisation, and mapping of different types of research tools:

As stated above, the WG will primarily extend the work of RDA groups working on global open research commons:

- Global Open Research Commons IG (GORC IG: Typology and Definitions)
  This RDA group is: (i) developing a shared understanding of what a 'commons' is within the research data space, (ii) connecting relevant national, regional and international initiatives; and, (ii) coordinating the delivery of a global Open Research Commons and monitoring related RDA groups.

---

[4] Hanisch, RJ; Kaiser, DL; Yuan, A; Medina-Smith, A; Carroll, BC; Campo, EM (2023) NIST Research Data Framework (RDaF) Version 1.5. (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 1500-18r1. https://doi.org/10.6028/NIST.SP.1500-18r1 (Accessed: 30 May 2023).

- **GORC International Model WG** (GORC WG Commons Attributes Model Version 0.5) This RDA group is: (i) generating a set of pertinent attributes to identify common features across open research commons by reviewing and identifying attributes or features currently implemented by a target set of GORC organisations and when possible identifying how they measure their user engagement with these features.

Other relevant Open Science/Research/Data commons initiatives:

- African Open Science Platform (AOSP) - A federated system that provides scientists and other societal actors with the means to find, deposit, manage, share and reuse data, software and metadata in pursuing their interests.

- China Science and Technology Cloud (CSTCloud) - A national platform to provide scientists with efficient and integrated cloud solutions in the retrieval, access, use, transaction, delivery and other aspects of sharing scientific information and relevant services.

- European Open Science Cloud (EOSC) - Contributes to the European Data Strategy by providing seamless access and reliable re-use of research data to European researchers, innovators, companies and citizens through a trusted and open distributed data environment and related services.

- Global Open Science Cloud - This initiative aims to encourage cooperation, and ultimately alignment and interoperability, between these and similar initiatives addressing the challenges of interoperability, technical infrastructure, policy and legal dimensions, and governance and sustainability.

- Malaysian Open Science Platform (MOSP) - A strategic transformative initiative to strengthen STI Collaborative Ecosystem for Malaysia that aims to make Malaysia's research data a valuable national asset by developing a trusted platform that enables accessibility and sharing of research data aligned to national priorities and international best practices.

Other relevant RDA groups:

- Data Granularity WG
- Data Repository Attributes WG
- Education and Training on Handling of Research Data IG
- FAIR for Virtual Research Environments IG
- FAIRsharing Registry: Connecting data policies, standards and databases RDA WG
- Global Open Research Commons IG
- GORC International Model WG
- Life Science Data Infrastructures IG
- Repository Platforms for Research Data IG

- [Research Data Architectures in Research Institutions IG](#)
- [Working with PIDS in Tools IG](#)
- [Research Data Architectures in Research Institutions IG](#)

Aggregators of research data tools:

- [RDM Training and Tools WG outcome](#) by El-Gebali S, Öjefors Stark K, Kronander, *et al*. ([SciLifeLab](#) RDM Training and Tools Working Group) - A Miro board identifying tools and services for open and reproducible research in the Life Sciences.
- [FAIRsharing](#) - A curated, informative and educational resource on data and metadata standards, interrelated to databases and data policies.
- [Re3data](#) - A global registry of research data repositories that covers research data repositories from different academic disciplines.
- [OpenDOAR](#) - A quality-assured, global Directory of Open Access Repositories.
- [COAR](#) - An international association that brings together individual repositories and repository networks in order to build capacity, align policies and practices, and act as a global voice for the repository community.
- [EOSC Marketplace and Portal](#) - Federation of services and tool related to Open Science, including aggregators, repositories, tools for the research lifecycle
- [OpenAIRE Graph](#) - An open resource that aggregates a collection of research data properties (metadata and links) available within the OpenAIRE Open Science infrastructure using a semantic graph database approach.

4. UN Sustainable Development Goals (SDGs)

Understanding the features, functionality, and interoperability of research tools within the global digital research data infrastructure landscape will help to support data management, sharing, and reuse to tackle grand societal challenges and address the United Nations Sustainable Development Goals (SDGs). In particular, this work contributes directly to SDG 17 which aims to 'Strengthen the means of implementation and revitalise the Global Partnership for Sustainable Development'.

5. Adoption Plan

This WG will undertake the necessary preliminary work for the creation of an online database of different types of research tools mapped to the stages of the research data lifecycle. This work aligns with the RDA's mission to build the social and technical infrastructure to enable researchers and innovators to openly share and re-use data across technologies, disciplines, and countries.

For transparent and accessible collaboration, the WG will use a [Google Folder](#) for its documentation. Updates will be regularly posted to the [WG wiki page](#) summarising meetings and sharing important updates relating to WG progress and timelines. The WG will organise regular dissemination activities and solicit community feedback during specific phases of the project. Community consultation (e.g., calls to action, surveys) may be employed to identify

different types of research tools used and required by community members throughout various stages of the research data life cycle. The WG will also collaborate with tool providers and Open Science/Research/Data Commons professionals to understand the fast-evolving digital research data landscape and ensure the WG deliverables meet the needs of adopters.

It will be important to validate WG deliverables (Section 1) with the global research data community (researchers, data support professionals, research tool developers/providers, research performing organisations, publishers, funders, and policymakers) at various stages of the WG's lifecycle.

The preliminary database of different types of research tools (Deliverable 3) is intended to be further developed to become a dynamic and community-curated resource in the future. As described above, the WG will develop recommendations for the long-term maintenance, sustainability, and adoption of the database by different stakeholders (outlined in Section 2).

6. Work Plan

A work plan has been defined that facilitates an efficient and timely delivery of WG deliverables. Working Group members will meet virtually via Zoom (for max. 90 mins) monthly from the end of May 2023. Tasks will be divided and allocated to task groups within the WG, and work undertaken by task groups in between meetings as required. Meetings will involve lightning updates from task groups and may include presentations from external speakers (e.g. tool providers, RDA groups, Open Science/Research/Data Commons professionals).

Month/Year	Preliminary Working Group Activities
April 2023	● First brainstorming workshop & publication of case statement ● Workshop slides, collaborative notes workshop 1 & workshop 2
May 2023	● Endorsement of case statement (Community, Council & TAB) ● 1st WG meeting (WG kick-off meeting & member consultation)
June 2023	● 2nd WG meeting (i. Presentation of WG aims, objectives, deliverables and timeline. ii. Allocation of task groups) ● Outreach (internal & external)
July 2023	● 3rd WG meeting (lightning update/working meeting/presentation) ● Outreach (internal & external)
August 2023	● 4th WG meeting (Deliverable 1: Creation of research data lifecycle framework and crosswalk to existing models) ● Outreach (internal & external)
September 2023	● 5th WG meeting (Allocation of task groups and preparation for P21 session) ● Definition of WG recommendations & outputs structure ● Review of RDA-OfR agreement (Internal)
October 2023	● 6th WG meeting at RDA's 21st Plenary Meeting in Salzburg (Present

	[WG progress and solicit feedback])	
	● Outreach (internal & external)	
November 2023	● 7th WG meeting ([lightning update/working meeting/presentation - collection/analysis of work from P21])	
	● Outreach (internal & external)	
December 2023	● 8th WG meeting ([Deliverable 2: The identification, categorisation, and mapping of different types of research tools])	
	● Outreach (internal & external)	
January 2024	● 9th WG meeting ([lightning update/working meeting/presentation])	
	● Outreach (internal & external)	
February 2024	● 10th WG meeting ([Deliverable 3: The creation of a preliminary structural framework for an online open access 'map of the digital research data tool landscape'])	
	● Outreach (internal & external)	
March 2024	● Final WG Recommendation Community review	
April 2024	● Final WG Recommendation Endorsement (Council) & Press campaign	

7. Initial Membership and Leadership

The WG will represent international perspectives from a variety of stakeholders, including researchers, data support professionals, system/service providers, policymakers, publishers, and librarians. Following two brainstorming workshops held in April 2023, the WG comprises the following initial membership and leadership*:

	Name	Affiliation	Country	Participation
1	Adam Leary	Oxford University Press	UK	Member
2	Adam Vials Moore	JISC	UK	Co-chair
3	Alex Moura	King Abdullah University of Science and Technology (KAUST)	Saudi Arabia	Member
4	Allyson Lister	FAIRsharing, University of Oxford	UK	Member
5	Christine Lemster	GEOMAR Helmholtz Centre for Ocean Research Kiel	Germany	Member
6	Cristiana Bettella	University of Padua	Italy	Member
7	Emmanuel Adamolekun	Helix Biogen Institute	Nigeria	Co-chair
8	Francis P. Crawley	CODATA International Data Policy Committee & EOSC-Future RDA Artificial Intelligence & Data Visitation Working Group	Belgium	Co-chair
9	Hea Lim Rhee	Korea Institute of Science and Technology Information (KISTI)	Korea	Co-chair
10	Kathryn Claypool	Arizona State University	USA	Member
11	Lauren Maxwell	University of Heidelberg, World Health Organization	Germany	Member
10	Lina Harper	Digital Research Alliance of Canada	Canada	Member

12	Lisa Curtin	figshare	USA	Member
13	Louise Bezuidenhout	DANS	Netherlands	Member
14	Luc Betbeder-Matibet	UNSW	Australia	Member
15	Maggie Hellström	ICOS Carbon Portal & Lund University	Sweden	Member
16	Malgorzata Lagisz	University of New South Wales Sydney	Australia	Member
17	Marcelo Garcia	King Abdullah University of Science and Technology (KAUST)	Saudi Arabia	Member
18	Marina Razmadze	Institute for Scientific and Technical Information	Georgia	Member
19	Meredith Goins	WDS-IPO	USA	Member
20	Natalie Meyers	Lucy Family Institute for Data & Society, University of Notre Dame	USA	Member
21	Nina Weisweiler	Helmholtz Association	Germany	Member
22	Noel Chibhira	University Of Pretoria	UAE	Member
23	Paolo Manghi	CNR-ISTI & OpenAIRE AMKE	Italy	Member
24	Rebecca Koskela	RDA-US	USA	Member
25	Richard Pitts	Oracle for Research	UK	Member
26	Rory Macneil	Research Space	UK	Co-chair
27	Ross Maxwell	Centre for In Vivo Imaging, Newcastle University	UK	Member
28	Sarah Stewart	University of Oxford	UK	Member
29	Shawna Sadler	ORCID	Canada	Member
30	Stefanie Kethers	ARDC	Australia	Member
31	Susanna-Assunta Sansone	University of Oxford, UK	UK	Member
32	Ville Tenhunen	EGI Foundation	Netherlands	Member
33	Xin Chen	Chinese Academy of Sciences	China	Member

*Upon endorsement, the WG aims to recruit members from Asia-Pacific countries (East Asia, South Asia, Southeast Asia, and Oceania).

Community comments addressed

Comment 1: National Institute of Standards and Technology

Author: Robert Hanisch

Date: 11 May 2023

The following is the NIST Team response to a review of the new RDA-OfR Working Group Case statement. We appreciate the opportunity to comment as the case proposal is reviewed and revised.

The case statement proposed by the RDA-OfR Working Group appears to overlap with the aims of the NIST Research Data Framework (RDaF).1

Comment addressed: The NIST RDaF uses a lifecycle approach with six high-level lifecycle stages, topics, and subtopics to organise key information concerning RDM and research data dissemination. The RDaF does not categorise or map different types of tools to the research data lifecycle nor does it address interoperability between tools.

The RDaF, which was released in Version 1 in February 2021,2 has just been updated to Version 1.5 and already accomplishes much of what appears to be proposed by this version of the Case Statement, although the Case Statement would benefit by a tighter focus.

Comment addressed: NIST RDaF version 1.5 has been cited in Section 3 of the case statement and will be used as one of several frameworks to create/select a harmonised research data lifecycle framework to categorise and map different types of research tools.

For example, the Case Statement notes that the "data infrastructure landscape comprises thousands of different systems, tools, and platforms for managing and sharing research data" but does not address the data itself. In succeeding sections, the Case Statement discusses data management and sharing systems. What does "data systems" mean for the group? Is it a catalog of infrastructural components such as repositories, AI platforms, and analysis software or something else? Good definitions would be helpful. The RDaF has definitions for many relevant concepts and terms.

Comment addressed: The term 'data systems' has been replaced by 'research tool' throughout the case statement to be clear. The term 'tool' can be defined as 'something tangible, such as a template or software program, used in performing an activity to produce a product or result', in accordance with CODATA's Research Data Management Terminology.

Clarifying the focus of the Case statement is important to ensure there is extension and not duplication of the work done on the RDaF being developed under NIST in the US with international input and reach. It would be appropriate for the Case Statement to cite the RDaF and recognize how this initiative relates to that one. From there, it is important to promote coordination and cooperation with the RDaF. We'd like to suggest a few points where clarification and cooperation is needed to avoid duplication and achieve synergy.

Comment addressed: The focus of the statement has been revised and clarified to ensure extension of and contribution to existing work in the area, including the use of NIST's RDaF to guide the identification, categorisation and mapping of research tools. As stated above, NIST RDaF version 1.5 has been cited in Section 3.

"1. Conducting a literature review of existing work"

The RDaF cites over 600 references relevant to the research data management (RDM) ecosystem, including best practices, policies, and vocabularies. These references can be linked to the "topics and subtopics" in the RDaF Framework Core which may or may not relate to components in the RDA Case Statement. The RDaF also catalogs more than 100 organizations, national and international, participating in some aspect of research data. These organizations are part of the data management landscape included in the RDaF.

"2. Creating an ontology and conceptual map of data management and sharing systems (Output 1)"

The Case Statement indicates it will look at all states of the lifecycle. The RDaF organizes the research data ecosystem into six major lifecycle stages, each of which has an extensive list of topics and subtopics. The relationships among these topics and subtopics are identified through 14 overarching themes and 8 professional "profiles" describing typical roles and responsibilities of people whose jobs influence or are influenced by RDM issues. The RDaF team explored both domain-specific and domain-agnostic approaches, finding that while there are unique issues in certain domains, the challenges of RDM largely transcend those specialties. It would be important to relate this work to what the Case Statement Team expects to accomplish and to use the RDaF work already done.

Comment addressed: As above. The co-chairs agree the NIST RDaF will be a useful resource to create/select a harmonised research data lifecycle framework to help with the categorisation and mapping of different types of research tools.

Additionally, how will the proposed catalog of data systems be distinct and extend what re3data and FAIRsharing.org curate?

Comment addressed: As outlined in Section 3 of the case statement, FAIRSharing, re3data.org and other data aggregator platforms/services will be researched and consulted to identify, categorise and map different types of research tools to the research data lifecycle. One important distinction is that the WG aims to categorise and map different *types* of research tools; it is beyond the scope of this group to create a registry or directory which aggregates an exhaustive list of individual research tools. Furthermore, the WG aims to categorise and map various different types of research tools throughout all stages of the research data lifecycle which extends beyond repositories and databases. Finally, the WG will address tool interoperability.

"3. Designing a preliminary framework for an online open access reference resource detailing different data management and sharing systems (Output 2)"

Depending on the definition of "… systems" in the Case Statement, this could build on what the RDaF has already done. The RDaF does not make recommendations about what systems or services should be used nor does it list specific systems, tools or technologies, although some generic elements are noted. This is intentional, as NIST is strictly neutral

when it comes to implementation technologies. Rather, it provides topics that research data organizations need to consider in making decisions concerning specific systems, services, and tools. If the Case Statement describes these and other infrastructure components that would be used in addressing RDaF topics, this might be a useful extension of the existing RDaF work.

Comment addressed: The WG aims to provide information about features and functionalities of different types of research tools, beyond generic definitions. The WG also intends to assess the interoperability of different types of tools used throughout stages of the research data lifecycle. This information will be a useful contribution to RDaF version 1.5.

Development of the RDaF has been a nearly four-year, $2M effort that was built on community engagement through three plenary workshops and 15 stakeholder workshops, involving more than 300 professionals from across the RDM spectrum. RDA-US contributed to this through in-kind support of a consultant. The RDaF Steering Committee is chaired by the former secretary general of CODATA and has three international members, including the secretary general of RDA and the president of CODATA. Given the familiarity of these individuals with the RDaF, we see opportunities to help revise the current draft Case Statement so that it leverages what has already been done and channels additional efforts to identify RDM infrastructure components that could be used in the implementation of the RDaF.

We would be pleased to see the RDA provide valued-added contributions to the RDaF, such as implementation support to organizations desiring to assess and improve their RDM capacity, and hope the Case Statement can be revised to create that synergy. We also note that the Australian Research Data Commons recently released its "RDM Framework for Institutions." While it appears to be primarily focused on Australian universities, there may also already be much information available therein that does not need to be repeated by the RDA. Similarly, FAIRSharing.org – which grew out of an RDA WG – already indexes numerous RDM service providers and is actively curating their metadata collection.

Comment addressed: The ARDC Research Data Management Framework for Institutions is an Australian national framework that features 19 essential elements for research data management. To our knowledge, this framework provides a guide to help institutions: (i) design RDM policy, procedures, infrastructure and services; and, (ii) improve coordination of RDM within and between institutions. It does not identify, categorise or map different types of research tools to the research data lifecycle. As outlined in Section 3 of the case statement, this resource will be used in conjunction with the NIST RDaF to create/select a harmonised research data lifecycle framework to guide the identification, categorisation, and mapping of different types of research tools.

Comment 2: FAIRSharing

Author: Susanna-Assunta Sansone

Date: 17 May, 2023

As a co-chair of the RDA FAIRsharing WG I also appreciate the opportunity to comment as this case proposal is reviewed and revised.

I agree with Robert's comments. It is essential to avoid duplication where considerable time and effort from organisations and stakeholders have gone into creating similar resources that are widely used and adopted.

Comment addressed: As stated above and as outlined in Section 3 of the case statement, FAIRSharing and other data aggregator platforms/services will be researched and consulted to identify, categorise and map different types of research tools to the research data lifecycle. One important distinction is that the WG aims to categorise and map different *types* of research tools; it is beyond the scope of this group to create a registry or directory which aggregates an exhaustive list of individual research tools. Furthermore, the WG aims to categorise and map various different types of research tools throughout all stages of the research data lifecycle which extends beyond repositories and databases.

There are many aggregators of data systems, and it is clear that users need help in finding resources throughout the various stages of the research data life cycle. However, the key challenge of such aggregators is that the broader (and ambitious) is their coverage, the shallower (and inaccurate) is their content, often failing to deliver reliable and trustworthy advice to the users. Among others, the success of such aggregators depend on their ability to: (i) strike a balance between content depth and breadth, (ii) map and harmonize information extracted from different sources, (iii) have their content community-vetted, and (iv) keep it up-to-date. These require continuous community input and contributions, as well as considerable time and effort, which goes well beyond the 18 months of a WG life span. Last but not least, a sustainability plan (and community support) is vital to mantain and grow it, keeping it relevant, and most importantly open, and freely available. It is not clear how these will be addressed by the proposed WG.

Comment addressed: Deliverable 3 of the WG aims to create a preliminary structural framework for an online open access 'map of the digital research data infrastructure landscape' that will require development and iterative updates by the community beyond the WG lifespan. To ensure the database remains current, relevant and useful to the wider research data community after the WG has completed its work, the WG must also provide recommendations for the long-term maintenance, sustainability and adoption of the database (Section 1).

Therefore, I strongly encourage the proponents to envisage also a way to provide valued-added contributions to existing primary resources, which already map (even if partially) the data infrastructure landscape. This is the case of the RDA-recommended FAIRsharing, interlinking standards, databases and policies across all disciplines. Being more focused, FAIRsharing is able to enrich, harmonize and curate the description of its

content also with the input of the FAIRsharing Community Champion Programme, partly supported by an RDA grant.

Comment addressed: FAIRSharing is one of several data aggregator services that will be researched and consulted by the WG to identify, categorise and map different types of research tools to the research data lifecycle. The WG will ensure appropriate acknowledgement of all existing work in the area used to produce its deliverables. The work of this WG adds-value to FAIRSharing, as a service provider, by: (i) improving understanding of the different types of research tools operating within the digital research data landscape; and, (ii) providing information about research tool features, functionalities, and interoperability to enhance data management and sharing practice.

Comment 3: FAIRSharing

Author: Allyson Lister and Susanna-Assunta Sansone
Date: 17 May, 2023

General Comments on
https://www.rd-alliance.org/group/rda-ofr-mapping-digital-research-data-infrastructure-landscape-wg/case-statement/rda-ofr
 by Allyson Lister and Susanna-Assunta Sansone

Thank you for providing the case statement for public review. We would like to agree with Bob Hanisch's comments, and also add a few more. We look forward to seeing where this new WG might go, especially within the context of already existing projects and WGs; it is in these collaborations that we are strongest.

Updates of this case statement should explicitly state an understanding of existing projects, WGs, and endorsed outputs such as FAIRsharing (beyond just mentioning them as in the current draft), and the *value* you expect to a) add to these resources (so as not to reinvent the wheel), and b) gain by creation of a new resource (mentioned in item 3). Such a new resource would have to be tightly scoped, if indeed one is needed at all. There is little interest among the research data community for the creation of 'yet another' repository. A WG such as this one should show how you would *contribute* to existing resources just as much, if not more, than creating new ones.

Comment addressed: Section 3 includes a list of existing work in the area accompanied by brief descriptions of relevant projects, initiatives and organisations where applicable. The WG scope and case statement content has been revised and refined to demonstrate the added value and novelty of this WG.

What follows are more specific comments related to the outputs listed in the case statement:

1. Conducting a literature review of existing work

A 'literature' review is not really an appropriate name for what could end up being a review of many tools, registries, databases and perhaps also publications. This sounds like a very large task, and perhaps should be constrained a bit more before finalising the case statement.

Comment addressed: In Section 1 (paragraph 4), the term 'literature review' has been replaced by 'landscape review'.

2. Creating an ontology and conceptual map of data management and sharing systems (Output 1)

We have already implemented two terminologies within FAIRsharing; one for research areas, and one for technical domains. Within the domain terminology, we are in the process of implementing a small sub-vocabulary for the types of content within databases. For this reason, and others stated in this comment, you should add the FAIRsharing WG to the "RDA WGs related to digital research data infrastructures" list to ensure that your work aligns with this WG.

Comment addressed: The FAIRsharing Registry: Connecting data policies, standards and databases RDA WG has been included in the list of relevant RDA groups to collaborate with and ensure work is aligned.

3. Designing a preliminary framework for an online open access reference resource detailing different data management and sharing systems (Output 2)

One output is "a catalogue different types of data systems" and that will provide "clear definitions to describe them based on their features and functionalities." Further, you describe 'data systems' as including: specialist research tools, data management planning tools, electronic laboratory notebooks, virtual research environments, databases, repositories, portals, archives, and data sharing platforms.

FAIRsharing (officially endorsed by the RDA and with its own FAIRsharing working group at (https://www.rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html) and other registries are already catalogues of many different resource types, including many of what you describe as 'data systems'. This means that not only is the 'framework' you propose (whether or not you implement such a registry) a massive undertaking; but it is also one that overlaps greatly with existing registries such as FAIRsharing or re3data, both very large projects in themselves. To take the example further, FAIRsharing describes a variety of databases (many of which fit within your definition of data systems) in curated form and linked to other resources (standards and policies) as part of an ecosystem of relationships. FAIRsharing contains over 1700 curated database descriptions (and nearly 4000 curated descriptions if you include standards and policy records), and more are being added all the time. Therefore, I suggest that this section be written more clearly, as the creation of a catalogue as you describe can only happen with the input of existing catalogues of data systems such as FAIRsharing and re3data, and building on what NIST

provides. It could even be that contributions to existing registries would be a much better use of resources and funding.

For example, one method of addressing this might be to have WG member(s) join the FAIRsharing Community Champion Programme to ensure that curation and educational content is added/updated there, to prevent duplication of efforts.

Comment addressed: As stated above, FAIRSharing and other data aggregator platforms/services will be researched and consulted to identify, categorise and map different types of research tools to the research data lifecycle. One important distinction is that the WG aims to categorise and map different types of research tools; it is beyond the scope of this group to create a registry or directory which aggregates an exhaustive list of individual research tools. Furthermore, the WG aims to categorise and map various different types of research tools throughout all stages of the research data lifecycle which extends beyond repositories and databases, and to address the under-examined issue of tool interoperability.