RDA-OfR Mapping the digital research data infrastructure landscape WG

Charter

The digital research data infrastructure landscape comprises thousands of different systems, tools, and platforms for managing and sharing research data during various stages of the research data lifecycle. Such data systems vary widely depending on data type, user requirement, system provider and subject area[1].

The diversity of data systems can prove overwhelming and challenging for stakeholders working within the research data ecosystem to navigate and select the most appropriate system to meet their needs and objectives. The variety of data systems gives rise to several challenges; definitions of related terms (such as 'data', 'dataset', 'database' and 'repository') and their functional differences remain unclear. In many cases, data systems are not interoperable, which leads to siloed working within organisations and disciplines, thereby limiting the scope of research and utility of research data.

This RDA Working Group (WG), supported by Oracle for Research (OfR), aims to define, map and enable stakeholders to navigate the complex digital research data infrastructure landscape by:

1. **Conducting a literature review of existing work**
   First, the WG will research and consult existing resources (registries/directories, organisations, projects, and RDA groups) to identify different types of data systems used throughout various stages of the research data life cycle (e.g., specialist research tools, data management planning tools, electronic laboratory notebooks, virtual research environments, databases, repositories, portals, archives, and data sharing platforms). This initial activity forms the basis for creation of a high-level ontology and conceptual map of the digital research data infrastructure landscape.

2. **Creating an ontology and conceptual map of data management and sharing systems (Output 1)**
   Next, the WG will catalogue different types of data systems and create clear definitions to describe them based on their features and functionalities. To refine the scope of this activity, the WG will focus on cataloguing domain-agnostic data systems.

3. **Designing a preliminary framework for an online open access reference resource detailing different data management and sharing systems (Output 2)**
   Based on the creation of the ontology and conceptual map, the WG will design the structural framework for a dynamic online reference resource, hosted by the RDA, owned by the community, and powered by Oracle for Software, that contains searchable information about different domain-agnostic data management and

---

[1] DARE UK Consortium. (2021). UK Data Research Infrastructure Landscape. https://doi.org/10.5281/zenodo.5584696

sharing systems. The preliminary framework provides the foundation for the development of an open access, crowdsourced, autonomous database that can be continuously populated by the community to incorporate domain- and region-specific data systems.

Value Proposition

Providing the global research data community with an up-to-date online database of data systems benefits all stakeholders working within the research data ecosystem. The outputs and recommendations produced by this WG aim to help:

Adopters	Value/Impact
Researchers (e.g., data creators and data users) …	Navigate, understand, and select suitable data systems 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) …	Provide relevant training and education for researchers thereby enabling them to select systems most suitable for their needs and objectives.
System developers/ providers…	(i) Understand the different systems operating within the digital research data landscape, and (ii) improve the features, functionalities, harmonisation and interoperability of their own data systems to enhance data management and sharing practice.
Research performing organisations…	Make informed recommendations at the institutional policy level to staff about appropriate systems for the management and sharing of research data.
Publishers…	Make informed recommendations to authors and journal editors about appropriate systems for the management, publication and sharing of data associated with journal manuscripts.
Funders…	Make informed recommendations to researchers and project managers based on data management plans for funded research.
Policymakers…	Make appropriate recommendations regarding the integration of data science and data policy into public policy. This should also include attention to changes in assessment practices for research, researchers, and research performing institutions.

Engagement with existing work in the area

This WG will conduct a literature review of existing work related to mapping the digital research data infrastructure landscape. This will require engagement with:

- Registries/directories of data management and sharing systems, including but not limited to FAIRsharing, OpenDOAR and Re3data.
- External organisations, including National Institutes of Health (NIH), National Science Foundation (NSF), National Information Standards Organization (NISO), International Organization for Standardization (ISO), UNESCO, International Science

Council, CODATA, World Data System (WDS), World Health Organization (WHO), European Open Science Cloud (EOSC) and Global Open Science Cloud (GOSC).
- International projects, including WorldFAIR and FAIR-IMPACT
- FAIR, CARE and TRUST principles
- RDA WGs related to digital research data infrastructures:
  - Data Granularity WG
  - Data Repository Attributes WG
  - FAIR for Virtual Research Environments 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

Adoption Plan

The RDA-OfR Mapping the digital research data infrastructure landscape WG will undertake the necessary preliminary work for the creation of an online database of data systems covering different disciplines and regions. This work aligns with the RDA's 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. The WG will organise regular dissemination about its activities and solicit community feedback during specific phases of the project. Community consultation (e.g., calls to action, surveys) will be required to identify different types of data systems used by community members throughout various stages of the research data life cycle across different disciplines and regions. To ensure the final output has global value, it will be critical to include data systems developed in countries where English is not the primary language.

Since the online database of data systems is intended to be a dynamic, living resource that evolves and adapts with the research data community over time, a citizen science approach will be taken to crowdsource information (data and metadata) about different data systems. The preliminary framework for the database must be circulated and validated by the research data community.

To facilitate its widespread adoption, RDA and OfR WG members will engage various stakeholders within the global research data community (researchers, data support professionals, system developers/providers, research performing organisations, publishers, funders, and policymakers), non-governmental and governmental organisations.

Work Plan

A work plan has been defined that facilitates an efficient and timely delivery of quality outputs and recommendations. 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 subgroups within the WG, and work undertaken by subgroups in between meetings as required.

Month/Year	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 (Networking, WG kick-off & literature review)
June 2023	● 2nd WG meeting (Report back on literature review: presentation of data sharing system types) ● Outreach (internal & external)
July 2023	● 3rd WG meeting (Ontology & conceptual mapping of data systems) ● Outreach (internal & external)
August 2023	● 4th WG meeting (Ontology & conceptual mapping of data systems) ● Outreach (internal & external)
September 2023	● 5th WG meeting (Pre-final version Output 1: Ontology & conceptual map) ● 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 (Final version Output 1: Ontology & conceptual map) ● Outreach (internal & external)
December 2023	● 8th WG meeting (Preliminary framework for online resource) ● Outreach (internal & external)
January 2024	● 9th WG meeting (Pre-final version Output 2: Preliminary framework for online resource) ● Outreach (internal & external)
February 2024	● 10th WG meeting (Final version Output 2: Preliminary framework for online resource) ● Outreach (internal & external)
March 2024	● Final WG Recommendation Community review
April 2024	● Final WG Recommendation Endorsement (Council) & Press campaign

Initial Membership and Leadership

The WG will represent international perspectives from a variety of stakeholders, including researchers, data support professionals, system/service providers, policy makers, 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	Hea Lim Rhee	Korea Institute of Science and Technology Information (KISTI)	Korea	Co-chair
2	Emmanuel Adamolekun	Helix Biogen Institute	Nigeria	Co-chair
3	Francis P. Crawley	CODATA International Data Policy Committee & EOSC-Future	Belgium	Co-chair

		RDA Artificial Intelligence & Data Visitation Working Group		
4	Rory Macneil	Research Space	UK	Co-chair
5	Adam Vials Moore	JISC	UK	Co-chair
6	Marcelo Garcia	King Abdullah University of Science and Technology (KAUST)	Saudi Arabia	Member
7	Alex Moura	King Abdullah University of Science and Technology (KAUST)	Saudi Arabia	Member
8	Ross Maxwell	Centre for In Vivo Imaging, Newcastle University	UK	Member
9	Natalie Meyers	Lucy Family Institute for Data & Society, University of Notre Dame	USA	Member
10	Rebecca Koskela	RDA-US	USA	Member
11	Maggie Hellström	ICOS Carbon Portal & Lund University	Sweden	Member
12	Kathryn Claypool	Arizona State University	USA	Member
14	Richard Pitts	Oracle for Research	UK	Member

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