# Case Statement: Data Repository Attributes Working Group

## 1. Charter

The Data Repository Attributes Working Group seeks to produce a list of common attributes that describe a research data repository and to provide examples of the current approaches that different data repositories are taking to express and expose these attributes.[1]

The working group will produce two documentary outputs over the course of 18 months and four Research Data Alliance (RDA) plenary meetings; they are:

1. a list of common descriptive attributes of a data repository with
    a. a definition of each attribute,
    b. a rationale for the use and value of each attribute,
    c. the feasibility of its implementation,
    d. a gap analysis of its current availability from data repositories, and

2. a selection of examples that illustrate the approaches currently being taken by repositories to express and expose these attributes to users and user agents.

The list of descriptive attributes of a research data repository will be submitted for review and endorsement to become an RDA Recommendation, and the selection of exemplars will be submitted for consideration as an RDA Supporting Output. This work is planned to take place over 18 months between January 1, 2022 and June 30, 2023.

## 2. Value Proposition

A complete and current description of a research data repository is important to help a user discover a repository; to understand the repository's purpose, policies, functionality, and other characteristics; and to evaluate the fitness for their use of the repository and the data that it stewards. Many repositories do not provide adequate descriptions in their websites, structured metadata, and documentation, which can make this challenging. Descriptive attributes may be expressed and exposed in different ways, making it difficult to compare repositories and to enable interoperability among

[1] This effort was catalyzed by the FAIRsharing WG session on "Repository Features Across Initiatives" at the 17th RDA Plenary Meeting.

repositories and other infrastructures such as registries. Incomplete and proprietary repository descriptions present challenges for stakeholders such as researchers, repository managers, repository developers, publishers, funders, and registries to enable the discovery and comparison of data repositories. For example:

- As a researcher, I would like to be able to generate a list of repositories to determine where I can deposit my data based on a query of descriptive attributes that are important to me.
- As a repository manager, I would like to know what attributes are important for me to provide to users in order to advertise my repository, its services, and its data collections.
- As a repository developer, I would like to know how to express and serialize these attributes as structured metadata for reuse by users and user agents in a manner that is integrated into the functionality of my repository software platform.
- As a publisher, I would like to inform journal editors and authors of what repositories are appropriate to deposit their datasets that are associated with manuscripts that are being submitted.
- As a funder, I would like to be able to recommend and monitor data repositories to be utilized in conjunction with public access plans and data management plans for the research that I am sponsoring.
- As a registry, I would like to be able to easily harvest and index attributes of data repositories to help users find the best repository for their purpose.

While this is not an exhaustive list of stakeholders and potential use cases, the value of identifying and harmonizing a list of descriptive attributes of data repositories and highlighting current approaches being taken by repositories would help the community address these important challenges and move towards developing a standard for the description and interoperability of information about data repositories. The statements of interest below demonstrate that there is a significant interest in this work.

### 3. Engagement With Existing Work in the Area

Many sets of attributes have been identified by different initiatives with differing scopes and motivations.[2] These attributes have included information about data repositories such as terms of deposit, subject classifications, geographic coverage, API and protocol support, funding models, governance, preservation services and policies, openness of the underlying infrastructure, adherence to relevant standards and certifications, and

[2] E.g., Metadata Schema for the Description of Research Data Repositories, Repository Features to Help Researchers: An invitation to a dialogue, Identifying ELIXIR Core Data Resources, Core Trust Seal, Science Europe, The TRUST Principles for digital repositories, COAR Community Framework for Good Practices in Repositories, NIH: Selecting a Repository for Data Resulting from NIH-Supported Research, OpenDOAR Repositories and Metadata Practices, DCAT

more. The results of these efforts reflect the variety of stakeholders and the diversity of repository attributes of interest across different communities. The harmonization of a common set of repository attributes, accompanied by the rationale for these attributes, will provide the community with a clearer understanding of the needs and requirements of different communities, and this commonality can enable greater interoperability across repositories, registries, and other data infrastructures.

**4. Work Plan**

The proposed co-chairs of the working group have submitted a Birds of a Feather (BoF) session proposal for RDA P18 to engage stakeholders in further discussion around these issues and revision of this case statement, if necessary. It will meet monthly via Zoom throughout the 18 months with a rotation of co-chairs formulating and sharing agendas in advance and leading each meeting. All meetings will be open to the community and progress towards the two deliverables will be noted on the RDA wiki. Correspondence between meetings will take place using an RDA mailing list that will be archived and accessible to all RDA members. The working group will strive to achieve consensus in its decision-making through open and respectful discussion. All views will be recorded from the deliberations of the group (e.g., mailing list archive, wiki) for consideration and community review of its outputs.

The working group will refine its methods based on feedback from the review of this case statement and member input, but in general we anticipate taking these actions:

1. Identify current standards and approaches to describing data repositories
2. Define use cases/user stories for utilizing metadata about data repositories
3. Draft a list of attributes and the rationale for their use/value
4. Conduct focus groups to validate/refine list
5. Perform an environmental scan to identify exemplars of different approaches
6. Submit a list of data repository attributes as an RDA Recommendation for broader community input, review, revision, and adoption
7. Submit a selection of exemplars as a RDA Supporting Output
8. Outreach to present and promote the adoption of the outputs

Milestones will include:

1. BoF session at RDA P18 and approval of case statement
2. Monthly meetings commencing January 2022
3. Identification of current descriptive approaches, use case definitions, and first draft of attributes (Action items 1-3) before RDA P19

4. Environmental scan to identify exemplars, completion and submission of list of attributes as RDA Recommendation three months after RDA P20
5. Revision of list based on community input and submission of exemplars as RDA Supporting Output by June 20, 2023
6. After conclusion of working group, presentation of outputs and report of early adoption at RDA P21

Minimally, the working group will engage the Metadata Interest Group (IG), Domain Repositories IG, and Repository Platforms for Research IG, and we will explore joint sessions as needed. Other interest and working groups as well as stakeholders from outside of the RDA will also be welcomed and encouraged to participate.

**5. Adoption Plan**

Primary adopters:

1. Repository managers
2. Repository software developers
3. Registries - e.g., FAIRsharing, OpenDOAR, re3data

Consultation and input from both the working group and the broader stakeholder community will be undertaken to identify

1. the relevance of the descriptive attributes drafted as the first output of the group, and
2. the feasibility of adopting these characteristics from the perspective of implementation.

Consultation on attribute relevance will allow an iterative process of development. Stakeholders will identify the important functionalities they need from repositories (such as facilitating the data peer review process for publishers and their authors, or integration with funder review processes); provide the rationale for why these characteristics are important for their community; and clearly articulate the aspects and functions needed to support their use cases.

This consultation will ensure that the final version of the list of descriptive attributes will both represent those repository attributes that are already in use as well as those which are of most relevance to our stakeholders, who will benefit directly from a harmonized, common list of attributes, and who will ultimately lead in their adoption and implementation.

**6. Initial Membership**

The working group will be led by co-chairs who represent international perspectives from a variety of stakeholders, including a variety of repositories, registries, publishers, and librarians.

Co-chairs:

- Matthew Cannon, Taylor & Francis (UK)
- Allyson Lister, FAIRsharing.org, University of Oxford (UK)
- Washington Segundo, Instituto Brasileiro de Informação em Ciência e Tecnologia (Brazil)
- Kathleen Shearer, Confederation of Open Access Repositories (Canada)
- Michael Witt, re3data, Purdue University (USA)
- Kazu Yamaji, National Informatics Institute (Japan)

Initial statements of interest have been expressed by:

1. Sjaeful Afandi, ISJD LIPI (Indonesian Institute of Sciences)
2. Erlin Arvelina, Politeknik Negeri Bandung
3. Ian Bruno, CCDC
4. Francis P. Crawley, GCPA & SIDCER
5. Joy Davidson, Digital Curation Centre, University of Glasgow
6. Scott Edmunds, GigaScience Press
7. Vinciane Gaillard, European University Association
8. Chris Graf, Wiley
9. Dasapta Erwin Irawan, RINarxiv and Institut Teknologi Bandung Sciences
10. Carsten Kettner, Beilstein-Institut
11. Danie Kinkade, BCO-DMO
12. Iryna Kuchma, EIFL
13. Hervé L'Hours, UK Data Service
14. Jennie Larkin, U.S. National Institutes of Health
15. Adam Leary, Oxford University Press
16. Mark Leggott, NDRIO / RDC
17. Thomas Lemberger, EMBO
18. Dawei Lin, U.S. National Institutes of Health
19. Catriona MacCallum, Hindawi
20. Maryann Martone, UCSD/INCF
21. Lautaro Matas, LA Referencia
22. Shanti Maulani, Dustira Hospital Nursing Academy
23. Dimitra Mavraki, Hellenic Centre for Marine Research
24. Kiera McNeice, Cambridge University Press
25. Peter McQuilton, GSK
26. Remedios Melero, Spanish National Research Council (CSIC)
27. Dagmar Meyer, European Research Council Executive Agency (ERCEA)

28. Amy Nurnberger, Massachusetts Institute of Technology
29. Megan O'Donnell, Iowa State University
30. Jonathan Petters, Virginia Tech
31. Olivier Pourret, UniLaSalle
32. Pedro Principe, University of Minho
33. Lisa Raymond, WHOI
34. Eloy Rodrigues, University of Minho
35. Keith Russell, Australian Research Data Commons (ARDC)
36. Shawna Sadler, ORCID
37. Jennifer Sanchez-Davies, JISC
38. Susanna-Assunta Sansone, University of Oxford
39. Sirarat Sarntivijai, ELIXIR
40. Natasha Simons, Australian Research Data Commons (ARDC)
41. Loekas Soesanto, Jenderal Soedirman University
42. Hendro Subagyo, RIN LIPI (Indonesian Institute of Sciences)
43. Shelley Stall, American Geophysical Union
44. Peter Suber, Harvard University
45. Marie Timmermann, Science Europe
46. Robert Ulrich, Karlsruhe Institute of Technology
47. Joris van Rossum, STM
48. Danielle Welter, University of Luxembourg
49. Sarala Wimalaratne, DataCite
50. Guillaume Wright, F1000 Research
51. Lesley Wyborn, Australian National University
52. Rika Yuliant, Telkom University Open Library