

Repository Platforms for Research Data: Case Statement for RDA working group

[Charter](#)

[Value Proposition](#)

[Engagement with Existing Work](#)

[Related Work](#)

[Work Plan](#)

[Deliverables](#)

[Matrix of Functional Requirements](#)

[Use Case and Functional Requirement Documentation](#)

[Milestones](#)

[Initial Use Cases](#)

[List of Functional Requirements](#)

[Mode and Frequency of Operation](#)

[Addressing Consensus and Conflicts](#)

[Community Engagement](#)

[Adoption Plan](#)

[Initial Membership](#)

[Co-Chairs](#)

[Members](#)

Charter

The RDA Repository Platforms for Research Data working group will gather and analyze research data use cases in the context of repository platform requirements. The primary deliverable will be a matrix relating use cases with functional requirements for repository platforms.

Value Proposition

As of the writing of this case statement, several communities in the research data realm somewhat “grope in the dark” when it comes to choosing, utilizing, deploying or developing the best possible repository platform to meet particular research data needs:

- Research institutions, including universities, often have an institutional repository that was set up perhaps a decade or more ago to primarily accommodate documents, such as dissertations, papers, or pre-prints of articles. They are now facing mounting demands for research data support, wonder whether or how their existing repository can meet these demands, and often have a difficult time identifying suitable

alternatives, be it as a replacement of or in addition to their existing repository (see some of the [Related Work](#) further below for examples).

- Developers of repository software may have created their product and its features for a particular set of use cases (for instance, for “Open Data” in government agencies) that they believe they understand well, but they cannot easily make their platform useful and known to other potential implementers with research data (such as in universities, companies, or applied research institutions).
- Researchers with very specific data use cases may be caught in the middle of the two aforementioned groups: their institutions may not know of alternatives to whatever repository platform they (may) already have, or be unwilling to explore them for lack of broader understanding of the alternatives’ functional capabilities; and the developers of repository platforms may be disinclined to address those use cases without knowing if they are more broadly applicable in the research data community. As a result, researchers may spend time, energy, and money away from their actual research to patch together repository(-like) solutions to their own data needs.

The primary target audience for the aforementioned matrix consists of developers and service providers of repository software. The functional requirements will influence the development of repository software and related services to better serve the use cases of the research data community.

Secondary audiences for the matrix may include:

- repository managers interested in identifying functional requirements that will satisfy specific use cases in their institution, consortium, or subject domain;
- grant applicants who need to satisfy funder requirements by writing a research data management plan, which very likely includes aspects related to repository platforms;
- funding agencies whose funding application reviewers may be aided by a listing of what may be expected from research data repository platforms; and
- researchers who may need to guide their institutions in selecting repository platforms for their emerging data needs.

Engagement with Existing Work

The proposed WG members are not aware of current work that would overlap with, or duplicate, the proposed activity. Following are links to work that focused on specific use cases, institutions, and/or repository platforms:

Related Work

- [Open data and the academy: an evaluation of CKAN for research data management](#)
- [Kaptur Technical Report](#)
- [Edinburgh DataShare – Reflections from a Data Repository Manager](#)
- [Dataverse Network Pilot at North Carolina State University](#)
- [SURA Reports Findings from Data Management Pilot](#)
- [How to Evaluate Options for a Research Data Catalogue & Repository](#)

Many of the proposed WG's members are expected to already be users, implementers, or developers of repository platforms for research data, or to be exploring such activities, so their awareness of ongoing related research in this area should feed proactively into the WG's ongoing work.

Work Plan

Deliverables

Matrix of Functional Requirements

This matrix will align a list of distilled research data use cases with functional requirements related to repository platforms. These requirements may also relate to specifications for a generic API, particularly in terms of interoperability and workflows. Each use case will tick a box under one or more functional requirements, indicating that the use case may be satisfied with a repository platform that supports the related functional requirement(s). The intersection boxes of use cases and functional requirements may also contain values expressing the importance or usefulness of the functionality for the use case, with details of that to be developed by the proposed WG.

Part of such a matrix *might*, for illustration purposes, look like this:

Functional requirements for repository -> Use cases I V	Assignment of DOIs to datasets	Interactive visualizations of datasets (charts, graphs, maps, etc.)	Displaying and exporting (to bibliographic software) citation for datasets
Research data related to a doctoral dissertation	2	3	1
Researcher data related to a journal article	1	2	1
Institutional research data - about institution, for broader audiences	4	1	3

Cell content legend: 1=essential, 2=desirable, 3=optional, 4=not important

Use Case and Functional Requirement Documentation

In support of the matrix, each use case and functional requirement will be described in more detail. Use cases will follow a narrative, user story format that “captures the 'who', 'what' and 'why' of a requirement in a simple, concise way”¹. This documentation will serve as a reference for anyone looking for more detail on a particular use case or functional requirement from the matrix.

The review of specific repository platform products vis-a-vis functional requirements or use cases is outside the scope of the proposed WG’s activities, but may be informed by its output.

Milestones

Initial Use Cases

One of the first milestones will be an initial list of use cases that is deemed to be representative of the broad needs of the research data community. This list will need to be edited, distilled, and analyzed in order to identify common themes and requirements.

List of Functional Requirements

A master list of functional requirements will be created as use cases are analyzed. This master list will be produced before combining the use cases and functional requirements in a matrix.

Mode and Frequency of Operation

The working group will primarily communicate asynchronously online using the mailing list functionality provided by RDA. Online voice meetings will be scheduled as needed; likely once per month. When possible, in-person meetings will also be scheduled; these will take place at RDA plenaries and at other conferences where a sufficient number of group members are in attendance.

Addressing Consensus and Conflicts

Group consensus will be achieved primarily through mailing list discussions, where opposing views will be openly discussed and debated amongst members of the group. If consensus cannot be achieved in this manner, the group co-chairs will make the final decision on how to proceed.

The co-chairs will keep the working group on track by setting milestones and reviewing progress relative to these targets. Similarly, scope will be maintained by tying milestones to specific dates, and ensuring that group work does not fall outside the bounds of the milestones or the scope of the working group.

¹ http://en.wikipedia.org/wiki/User_story

Community Engagement

The working group case statement will be disseminated to mailing lists in related communities of practice related to research data, in an effort to cast a wide net and attract a diverse, multi-disciplinary membership. Group activities, where appropriate, will also be published to related mailing lists and online forums to encourage broad community participation.

Adoption Plan

The documented matrix of use cases for research data and functional requirements for repository platforms will be disseminated to members of the primary and secondary audiences mentioned under the [Value Proposition](#) via the methods mentioned in [Community Engagement](#). They will be asked to provide feedback to the working group on what specific use cases the WG's output helped them to address, how useful the documented matrix was to them, and where additions or improvements to it could be made in future iterations. This feedback will determine whether a follow-up WG should be established or the current WG's efforts continued, depending on the RDA governance guidance for that at the time.

Initial Membership

Co-Chairs

- David Wilcox, DuraSpace <dwilcox@duraspace.org>
- Stefan Kramer, American University Library <skramer@american.edu>

Members

The membership of the working group will consist of the following individuals:

- André Schaaff, Observatoire astronomique de Strasbourg, <andre.schaaff@astro.unistra.fr>
- Amy L. Nurnberger, Columbia University, <anurnberger@columbia.edu>
- Suntae Kim, Korea Institute of Science and Technology Information, <kistiman@gmail.com>
- Mark Hahnel, FigShare, <mark@figshare.com>
- Angus Whyte, Digital Curation Centre, University of Edinburgh, <a.whyte@ed.ac.uk>
- Julie Fukuyama, National Diet Library, Japan, <ightic@gmail.com>
- Thomas Zastrow, Rechenzentrum Garching (RZG) der Max-Planck-Gesellschaft / MPI für Plasmaphysik, <thomas.zastrow@rzg.mpg.de>
- Laura Molloy, HATII, <laura.molloy@glasgow.ac.uk>
- Thomas Jejkal, KIT (IPE), <thomas.jejkal@kit.edu>
- Ralph Müller-Pfefferkorn, Technische Universität Dresden, <ralph.mueller-pfefferkorn@tu-dresden.de>

who will, together with the co-chairs, bring capacities from, and represent, the following communities of potential adopters and implementers:

- Academic and national government libraries
 - Research, science, and technology institutions
 - Repository platform developers
- from Europe, Asia, and North America.