

## **Case Statement for the [Archives and Records Professionals for Research Data IG](#) to be renamed the “Collections as Data IG”**

### **Introduction**

This group is aimed at collections professionals such as archivists, librarians, records managers and museum curators, as well as related professions such as IT professionals, knowledge scientists, and those involved in standards development, who serve in a range of critical roles: as experts in ensuring access, preservation, and reuse of digital records, objects, data, and collections; as provocateurs for good collections curation practices; and as advocates for the construction of responsible and sustainable infrastructures for information sharing. As articulated in the [Vancouver Statement on Collections as Data](#), “this [collections] stewardship role only grows in importance as artificial intelligence applications, trained on vast amounts of data, including collections as data, impact our lives ever more pervasively.” We recognise that there is increasing pressure on memory institutions to establish good models for responsible and culturally sensitive development of data resources while also navigating the challenges and opportunities provided by new technologies (Vancouver Statement, Principles 2, 4, 9-11), and this is not work that can be easily done in silos. This group seeks to provide a space for examining alignments in values, policy, and practice in collections work, broadly imagined, encouraging a vibrant exchange of expertise across collecting areas and domains of practice.

The proposed IG is explicitly aligned with RDA’s goal to build social bridges that enable the open and FAIR sharing of data. Specifically, the IG will build relationships among professionals who are committed to developing data sharing frameworks and practices that are able to withstand the various challenges introduced by the passage of time: technological obsolescence, loss of contextual understanding of data, and resource constraints that make it impractical to commit to preserving all data forever. By creating a space for collections professionals and data curators to come together, this IG has the potential to act as a launching pad for RDA Working Groups that will produce recommendations addressing these fundamental challenges to research data sharing, specifically by bringing preservation, arrangement and description, and appraisal methodologies from the cultural heritage and information management sectors to the wider RDA community. This IG also aims to bring the skills and competencies which have long existed in memory institutions into the wider RDA community, many of whom will not be aware of this existing expertise.

### **Objectives**

- To provide a forum for discussion on facilitating the responsible computational reuse of collections as data in FAIR, Open, ethical and transparent ways.
- To communicate and coordinate with collections professionals who may be outside of RDA to foster synergies, bring them into RDA, and bring the IG and relevant WG activities to the attention of the field.
- To advocate for collections professionals within the RDA community.

- To build dialogue between evolving RDM and Open Science professional roles (e.g. data stewards, data engineers) by advocating for recognition of skills and competencies already well-developed within the collections field.
- To develop workflows to support collections as data, such as licensing, citations, and data documentation, additionally responding to and actioning the recommendations outlined in the [WorldFAIR Project](#).

### ***Related Groups***

This group will complement the activities of existing groups such as the:

- [Libraries for Research Data IG](#)
- [Education and Training on Handling of Research Data IG](#)
- [Professionalising Data Stewardship IG](#)
- [International Indigenous Data Sovereignty IG](#)

### ***Participation***

The group is aimed at collection professionals and information stewards, who may additionally have responsibility for research data management in their organisation. We intend to engage with this community via their specific professional communication platforms, for example archives- or research data curation-specific listservs.

We also intend to facilitate exchange between professionals working in the cultural heritage and information sectors and the broader research data sharing and preservation communities, through the communication of best practices, facilitating discussions (e.g. webinars, workshops, and meetings), and the provision of practical workflows aimed at different community needs which support these best practices.

### ***Outcomes***

- Increased participation in RDA by collections professionals
- Practical recommendations to the RDA community for addressing at least the following issues:
  - workflows for collections as data
  - strategies for improving ethical access to research data collections
  - education in digital preservation strategies
- Increased impact of RDA outputs via their effective amplification across professional networks for cultural heritage and information science

### ***Mechanism***

The IG will use the RDA-provided email list to share news and generate conversation, and will connect where relevant with existing professional mailing lists where already in existence. In addition, we will meet via teleconference on at least a quarterly basis to discuss projects and report on relevant initiatives taking place at our own institutions.

## ***Timeline***

- First meeting of the IG as the Archives and Records Professionals for Research Data IG in Fall 2015
- IG last active at a Plenary meeting in Fall 2021
- RDA Plenary BoF for Collections as Data group in Fall 2023 - *Group re-formation*
- In person RDA Plenary Fall 2024

## ***Engagement***

The IG will continue to increase group membership via the Archives and Records Association (UK and Ireland), JISC Research Data Management Group, Society of American Archivists, etc. In addition, the IG has identified members who will act as liaisons to key groups (see table below).

| <b>Organisation Name</b> | <b>Cross Member</b> |
|--------------------------|---------------------|
| CODATA                   | Laura Molloy        |
| CENL Dialogue Forum      | Sally Chambers      |
| RDA FAIRsharing WG       | Allyson Lister      |
| DARIAH RDM WG            | Beth Knazook        |
| Data Worthy Collective   | Margaret Warren     |

## ***Activities***

At the Salzburg Plenary, the IG identified several themes on which to focus our attention:

### Themes

- 1) Collaboration
  - Potential outcome: sharing of theoretical and practical foundations for enabling creative and scholarly reuse of cultural records. Although this group is not exclusive to professionals working in memory institutions, a significant purpose of this group will be to steer consensus between those engaged in collections as data work, whose training encompasses data appraisal, the creation of collections, and metadata generation as well as preservation, with researchers, data stewards and other research support professionals, digital preservationists, IT and software developers, etc. engaged in complementary work.
- 2) Responsibility
  - Potential outcome: agreement and improved messaging around licensing for cultural heritage records as research data assets, with a recommendation on rights & licensing best practices; methods for credit creation and modification of collections records which better integrate researcher contributions; advice on responsible application of CARE principles to collections reuse; advocacy for ethical use of Artificial Intelligence in collections work.

- 3) Outreach
  - Potential outcome: inclusion of a diversity of organisations in the conversation, not just those that are well-funded at the national level. Development of approaches that serve the wider profession and research communities.
- 4) Alignment
  - Potential outcome: promoting agreement around both values and technical standards. De-siloing of the collections field is needed to support the dialogue between various organisations in delivering collections as data, which will require practical tools and workflows that demonstrate best practice and help to narrow the approaches taken by diverse institutions in providing access to collections in ways that also supports computational reuse. Recognition of the shared values between collections professionals and the research data field broadly.

### **Potential Group Members**

See list of 200+ group members on the former Archives and Records Professionals IG page:

<https://rd-alliance.org/groups/archives-records-professionals-for-research-data.html>

### **APPENDIX: Existing Contributions**

The proposed IG will strive to build upon and expand work that has already begun as part of the wider Collections as Data movement. Examples include:

1. Vancouver Statement on Collections as Data:  
<https://collectionsasdata.github.io/part2whole/vancouver/> (available in English, Spanish, French, Arabic, and also Dutch soon)
2. University of North Carolina Chapel Hill: “On the Books: Jim Crow and algorithms of resistance”. <https://onthebooks.lib.unc.edu/>
3. [Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper](#)
4. Europeana is data steward for the European Cultural Heritage Dataspaces
5. [A workflow to publish Collections as Data](#) (SSHOC Open Marketplace) and the “Checklist to Publish Collections as Data in GLAM Institutions”
6. DiSSCo uses FAIR Digital Objects <https://sandbox.dissco.tech/>
7. University of Illinois Archives Cybernetics Thought Collective (Digitized Surrogates)  
<https://digital.library.illinois.edu/collections/38ec6eb0-18c3-0135-242c-0050569601ca-1>
8. [WorldFAIR Project Cultural Heritage Image Sharing Recommendations](#)
9. [Datashets for Digital Cultural Heritage Datasets](#)
10. [An approach to assess the quality of data generated from GLAM collections using Jupyter Notebooks](#)
11. [Open a GLAM Lab](#)