

## 23rd Plenary Meeting Collaborative Notes

**Group(s) name(s) organising the session:** [Collections as Data IG](#)

**Session link:** [Gathering Metrics and Setting Boundaries: Reusing Collections as Data and the impacts of AI / Recopilando métricas y estableciendo límites: Reutilización de colecciones como datos y el impacto de la Inteligencia Artificial](#)

**Session scheduled date/time/breakout session:** Thursday, 14 November, Breakout 6, 10:00 – 11:30 CST

**Venue:** Plaza de La Autonomía, Aula Magna

### Session summary (for Group co-chairs)

We will use the content in the table below to highlight your work to the RDA community as a report organised by the Technical Advisory Board and to a wider audience through English & Spanish social media mentions.

Please complete ALL fields below by **Friday, 29th November, close of business** to be included in the report & social media activities.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Summarise the session in three sentences:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| The session profiled a number of efforts within the cultural heritage sector to support the reuse of digital collections as data, including work by the International GLAM Labs Community and the British Library Labs, 2 projects to model ethical and responsible reuse (Responsible AI and D-CRAFT) as well as an example of AI applied to abstracts of multilingual catalogue records. The goal of the session was to gather audience contributions on further initiatives and best practices used to foster best practices at other institutions and organisations.                                                                                                                                                                                                         |
| <b>Key outcomes/actions/takeaways</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <ol style="list-style-type: none"> <li>1. <i>There were not many examples of published AI policies and only about half the room raised their hands when asked about whether they thought AI policies were necessary.</i></li> <li>2. <i>This prompted a related conversation about the role that GLAM organisations play in 'gatekeeping' risk assessment so that decision-making around inappropriate reuse is not passed on to the users, and how to document the investment in that work.</i></li> <li>3. <i>The D-CRAFT toolkit is an under-utilised resource that could benefit from wider community uptake (or at least, more widely documented uptake!) Is there the potential for this IG to work on developing a machine-readable D-CRAFT analysis tool?</i></li> </ol> |
| <b>Synergies and/or possible collaborations identified with RDA groups and other groups:</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| <b>Highlight text that will be used in social media mentions (please make sure the text is clear and appropriate for public consumption &amp; comprehension)</b>                                                                                                                                                                                                                                                                                               |
| There are a growing number of resources available to help collections professionals reconceptualise collections as data, but still much work to be done to explore the ethical responsibility of both creators and users, which may differ from the research sector due to the unique nature of cultural heritage resources as objects of community memory as well as the complex role of collections professionals as both stewards and creators of datasets. |
| <b>Direct link to Group home page</b>                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <a href="https://www.rd-alliance.org/groups/collections-as-data-ig/">https://www.rd-alliance.org/groups/collections-as-data-ig/</a>                                                                                                                                                                                                                                                                                                                            |

Get involved in [RDA Community](#)  
 Check out [P23 programme sessions](#)

This meeting will take place according to the [RDA Code of Conduct](#)

## Attendee Check-in

Please complete this table to indicate your attendance (add rows as needed):

| Full Name                | Affiliation                                          | Location      | Email                    |
|--------------------------|------------------------------------------------------|---------------|--------------------------|
| Beth Knazook             | Digital Repository of Ireland                        | Ireland       | b.knazook@ria.ie         |
| Ricardo Casate Fernández | Ministry of Science, Technology and Environment      | Cuba          | rcasate@gmail.com        |
| Claudia Cordova Yamauchi | Consejo Nacional de Ciencia, Tecnología e Innovación | Perú          | ccordova@concytec.gob.pe |
| Sophie Miller            | Space Telescope Science Institute                    | United States | smiller@stsci.edu        |

## Collaborative Session Notes *(To be used by participants and chairs during the session)*

### Meeting presenters:

**Yasmeen Shorish**, Director of Scholarly Communications Strategies & Special Advisor to the Dean for Equity Initiatives, James Madison University (Session chair)

**Hannah Scates Kettler**, Associate University Librarian, Iowa State University (Session chair)

**Gustavo Candela**, Lecturer in Computer Science at the University of Alicante

**Santi A. Thompson**, Interim Associate Dean for Organizational Development, Learning and Talent,  
University of Houston

**Ayla Stein Kenfield**

**Mahendra Mahey** Senior Research and Development Adviser Tallinn University, Researcher at the  
Estonian National Museum and PhD Student at the University of Strathclyde

**Gimena del Rio Riande**, Researcher at CONICET/Universidad del Salvador

**Responsible AI project** - developing a template, model, artifact (name forthcoming) for articulating values

<https://www.lib.montana.edu/responsible-ai/>

<https://osf.io/yue9s>

**International GLAM labs** - many institutions are currently setting up a lab to collaborate and promote responsible utilisation

<https://glamlabs.io/>

<https://glamlabs.io/member-map>

Shared best practices and guidelines. 16 members got together to publish a book of best practices

<https://doi.org/10.1002/asi.24835>

Checklist for publishing collections as data: <https://arxiv.org/abs/2304.02603>

GLAM Labs workshop in Iceland: <https://github.com/hibernator11/workshop-notebooks-dhnb202>

Join GLAM Labs

<https://glamlabs.io/slack-channel-signup/>

<https://www.jiscmail.ac.uk/GLAMLABS>

[https://twitter.com/GLAM\\_labs](https://twitter.com/GLAM_labs)

## D-CRAFT

Open Access Assessment Toolkit for measuring the reuse of digital assets in the digital library field  
Project concluded in 2023

Part of the Digital Library Federation Assessment group

<https://reuse.diglib.org/toolkit/>

Ethical guidelines produced by the project - description of values and practical considerations when assessing reuse (available in both Spanish and Mandarin translation)

<https://reuse.diglib.org/ethical-guidelines/>

## Evolving GLAM Labs

British Library Labs (first lab in a national library aimed at reusers and remixing of library content)

Early on working with computational use of the data: AI curated books, "The Mechanical Curator"

<https://mechanicalcurator.tumblr.com/>

<https://www.flickr.com/photos/britishlibrary/albums/72157638544764936/>

Breakdown barriers and open to change, failure and experimentation

2015 Collections as Data portal launched at BL

26 reuse and remix cases already conducted as part of PhD research project

Looking for uses of digital cultural heritage that have led to social impact - invitation to the audience!

GLAM Labs TPDL presentation: <https://github.com/hibernator11/wikidata-review>

**BIREDIAL ISTE:** <https://biredial.istec.org/>

Automatic language detection use case with polyglot library, where abstracts do not have the language tagged

95% agreement when catalogued by humans

Errors from humans and language models as well,

Conclusion that human supervision needed, intervention of human taggers, required the development of an interaction tool

Consultation chats required from users, not ready for the public yet

### Q&A

Is there more available about the GLAM labs community?

What are the services in the laboratories, and products that a lab might offer? The website has different sections with computational models, for instance based on notebooks or how to publish collections as data. The Iceland project focused on guidance for how to design a collection from scratch. There are accompanying research articles based on the GLAM lab community - it's an idea generator. There aren't any services online, more like a display of services around the labs.

Is the library language AI work available?

Question for Santi & Ayla: are you aware of organisations that are currently using D-CRAFT and/or is there a feedback mechanism that could iterate and expand the toolkit in light of new AI uses?

Not aware currently. Would love to know if people are using it!

There's a contact form on the website to leave feedback including instances of how they were able to use the toolkit for their assessment practices.

Deliverable has a template takedown policy that came out of concerns from Indigenous communities. No formal follow-up.

D-CRAFT is not wedded to a particular research methodological approach or practice.

Question to the audience: What initiatives and best practices are used to foster best practices and your institutions and organisations?

What brought you to the session today?

Academic patrimony collection from different kinds of knowledge fields - seeing trends in AI influencing capacity to learn about data. Universities have to learn more about substantive actions in using the database of academic activity.

Research article on emerging and trending topics: <https://doi.org/10.1002/asi.24583>

How can we encourage a provenance chain within AI models to know that AI was used? Repositories are facing similar challenges.

Considerations of collections as data versus datasets?

Collections as data is often derived from physical objects in cultural heritage organisations. In the process of digitisation, you are transforming it into a computationally accessible object. A thing that could make it different from a dataset, they are usually about people, society more specifically identifiable.

Indigenous data governance comes up a lot with ocean and environmental data. Concern about the lack of ability to constrain AI.

Federally funded assets have no restrictions in place for AI.

Users have to opt-in to use of the data; in cultural heritage stewards you're placing the information to users, what do you need to present to the user for making that context known? Tension between ownership-stewardship and responsibility of setting up parameters of the use

Who is actually using Cultural Heritage?

BL labs keen to capture uses - funding was reliant on metrics on how the lab was enabling access, people were engaging with content, how machines were accessing content

Tracking AI use; not assuming all AI is commercial or nefarious

AI perhaps just enforcing the idea that our policies are not adequate to account for commercial or unwanted uses; e.g. humans and machines can do harm to humans

E.g. policy for AI use: <https://data.nls.uk/projects/ai-statement/>,  
<https://www.nlb.gov.sg/main/site/sure-elevated/for-the-public/generative-ai>

Current modes of identifying AI uses is tagging robot use by IT and facilitating blocking if necessary.

One of the points of D-CRAFT was investigating who was using materials and why? How to document investment on the work that has been done for CH organisations digitising materials? This conversation surfaces how little awareness we have around what ultimately happens with these materials. Need a more intentional commitment in responding to this question. Identifying the harm that can come from oblivious use of cultural materials, and the inverse (the good).

An interesting interview (Mahendra) did was around GLAMs have responsibility to assess the risk of releasing content online, so that they don't pass this on to users.

Do we need persistent identifiers for creators of AI models?

Impact - Capturing the stories of use tell us what is happening with the data

IDEA DCRAFT Phase 2:: It might be interesting to provide a machine-readable D-CRAFT analysis version of the impact and use of the data? (Gustavo: just brainstorming....never thought about measuring the impact ...always focused on reuse...)

True lack of attribution that occurs across the web undercut a lot of basic citation processes - breaks down pretty quickly when trying to assess over iterations over the web

Participants are invited to share:

**a. How institutions are tracking and encouraging reuse of collections as data: What kinds of metrics do we expect or want to track?**

**b. How institutions are writing terms of use, with respect to collections-as-data: Who are the stakeholders involved in crafting these terms? What are the expectations implied in these terms around reuse/reusability?**

c. In the group discussion/outcome: What informs an institution's stance on AI use of collections, as data?