2022 Webinar Series
Highlighting RDA Outputs

Hosted by RDA-US

RDA CURE-FAIR WG WEBINAR

Introducing 10 Things for Curating Reproducible and FAIR Research

September 22, 2022
Individual Members

12,784 members from 148 countries

69% Academic & Research
14% Public Administration
11% Enterprise & Industry
6% Other

RDA Groups

33 Working Groups
53 Interest Groups
1 Community of Practice
## 30 National Groups

<table>
<thead>
<tr>
<th>Asia</th>
<th>Czech Republic</th>
<th>Italy</th>
<th>Slovenia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Denmark</td>
<td>Lithuania</td>
<td>So Eastern Europe</td>
</tr>
<tr>
<td>Austria</td>
<td>Estonia</td>
<td>Hungary</td>
<td>Spain</td>
</tr>
<tr>
<td>Brazil</td>
<td>Europe</td>
<td>Ireland</td>
<td>Sweden</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>Finland</td>
<td>Netherlands</td>
<td>Portugal</td>
</tr>
<tr>
<td>Canada</td>
<td>France</td>
<td>Norway</td>
<td>Romania</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>Germany</td>
<td>Portugal</td>
<td>Bosnia and Herzegovina</td>
</tr>
<tr>
<td>Croatia</td>
<td>Greece</td>
<td>Region of the Americas</td>
<td>Region of the Americas</td>
</tr>
</tbody>
</table>

## 4 Contributing Regions

- Australia
- EU
- France
- United States

[www.rd-alliance.org/groups/rda-us](www.rd-alliance.org/groups/rda-us)

www.rd-alliance.org/groups/national-groups
Translation
Please note, this event is being translated from English to French. To hear this webinar in French, select the Globe on the control area.

Questions and Answers
1. Raise hand and ask question
2. Post to Q & A section of Zoom
3. Post question in Chat

Recording
Please note, this event is being recorded, and will be available on the RDA website.

Translation
Please note, this event is being translated from English to French. To hear this webinar in French, select the Globe on the control area.

Questions and Answers
1. Raise hand and ask question
2. Post to Q & A section of Zoom
3. Post question in Chat

Recording
Please note, this event is being recorded, and will be available on the RDA website.
2022 Webinar Series
Highlighting RDA Outputs

Hosted by RDA-US

Introducing 10 Things for Curating Reproducible and FAIR Research Webinar

Limor Peer  Mandy Gooch  Tom Honeyman  Thu-Mai Christian  Florio Arguillas
Thanks to....

Presenters and IG
Webinar Support - Meghan and Rebecca
French Interpreters - Digital Research Alliance in Canada
Spanish Interpreters - LA Referencia

www.rd-alliance.org/groups/rda-us
Upcoming Webinars in this 2022 Series

- Research Metadata Schemas WG - October 25 at 20:00 UTC
- DMP Common Standards WG - November 22 at 17:00 UTC

www.rd-alliance.org/groups/rda-us
RDA PODCAST SERIES
Hosted by RDA-US

Data Streams is a collection of conversations among members in the Research Data Alliance (RDA) community.

#1 May 2022 The Data Citation WG - A Working Groups Journey  Mark Parsons and Andreas Rauber
#2 June 2022 Data Repository Attributes Working Group Discussion  Matthew Cannon and Michael Witt
#3 July 2022 Data Granularity Working Group  Reyna Jenkyns, Brigitte Mathiak, Katy McNeill, Guangyuan Sun
#4 Sept 2022 Sensitive Data Interest Group  Nichola Burton, Aleks Michalewicz and Kristal Spreadborough
#5 Oct 2022 National PID Strategies WG

Subscribe: www.rd-alliance.org/podcast
Research Data Alliance 10th Anniversary Plenary meeting
Gothenburg, Sweden, March 21-23, 2023
2022 Webinar Series
Highlighting RDA Outputs

Thank You For Joining Us!
CURE-FAIR WG webinar

Florio Arguillas, Thu-Mai Christian, Mandy Gooch, Tom Honeyman, Limor Peer

Thursday, 22 September 2022
<table>
<thead>
<tr>
<th>Time</th>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 minutes</td>
<td>Introduction of RDA and RDA-US/Stephanie Hagstrom</td>
</tr>
<tr>
<td>2 minutes</td>
<td>Introduction of Webinar and Presenters/Tom Honeyman</td>
</tr>
<tr>
<td>40 minutes</td>
<td>CURE-FAIR WG presentation</td>
</tr>
<tr>
<td></td>
<td>• About the CURE-FAIR WG: Objectives, case statement, supporting outputs/Limor Peer</td>
</tr>
<tr>
<td></td>
<td>• 10 Things for Curating Reproducible and FAIR Research/Mandy Gooch</td>
</tr>
<tr>
<td></td>
<td>• Adopter stories/Florio Arguillas</td>
</tr>
<tr>
<td></td>
<td>• Using 10 CURE-FAIR Things/Thu-Mai Christian</td>
</tr>
<tr>
<td></td>
<td>• Future of 10 CURE-FAIR Things/Tom Honeyman</td>
</tr>
<tr>
<td>10 minutes</td>
<td>Q&amp;A: Moderator/Tom Honeyman</td>
</tr>
<tr>
<td>2 minutes</td>
<td>Closing/Stephanie Hagstrom</td>
</tr>
</tbody>
</table>
Introductions

- Limor Peer, Associate Director for Research & Strategic Initiatives, Institution for Social and Policy Studies (ISPS), Yale University
- Mandy Gooch, Research Data Archivist, Odum Institute, University of North Carolina at Chapel Hill
- Florio Arguillas, Research Associate at the Cornell Centre for Social Sciences (CCSS)
- Thu-Mai Christian, Assistant Director for Archives, Odum Institute, University of North Carolina at Chapel Hill
- Tom Honeyman, ARDC Research Software Program Manager, Australian Research Data Commons
Summary of the Problem

- A computational reproducibility standard has not been adopted as an integral part of normative scientific practice.
- There is a lack of a standardized approach to curating for reproducible and FAIR research.
Definitions

▪ **Computational Reproducibility**: The ability to recreate computational results from the data and code used by the original researcher.

▪ **Curation**: The management and preservation of digital data over the long-term; involves a set of actions, guided by FAIR principles, to determine what research outputs are worthy of sharing and preserving, why, for whom, by whom, and for how long.

▪ **Research Compendium**: The collection of the research artifacts necessary to independently understand and repeat the entirety of the analysis workflow from data processing and transformation to producing results. Also referred to as a “reproducible file bundle.”

▪ **Curating for Reproducibility (CURE-FAIR)**: The object of the curation is the research compendium; the goal of curating for reproducibility is to enable continued access to the compendium and its component parts which are necessary to independently reproduce the associated results.
Researchers, curators, publishers, funders, repository managers, reviewers, and others

Stakeholders can benefit from more clarity about:

- Whether computational reproducibility can be achieved (using the objects contained within a research compendium);
- The quality of the research compendium and its component parts (e.g., does it meet community standards for FAIR and for long-term archival preservation?)
CURE-FAIR WG Supporting Outputs

- **CURE-FAIR Challenges**: Describe the challenges of preparing and reusing materials required for computational reproducibility; collect information from various stakeholders about their challenges. DOI: [10.15497/RDA00063](http://10.15497/RDA00063)

- **CURE-FAIR Annotated Bibliography**: Provide a broader understanding of what it means to curate research artifacts (e.g., data, code, software) for the purposes of supporting research reproducibility. [https://www.zotero.org/groups/2868459/rda_cure-fair_subgroup_1/library](https://www.zotero.org/groups/2868459/rda_cure-fair_subgroup_1/library)

- **CURE-FAIR Practitioners**: Identify organizations/groups that have fully implemented CURE-FAIR workflows and learn about the various ways researchers and research-supporting organizations have implemented data curation tools, services, and/or workflows that support computational reproducibility; develop a standard form to collect profile information from CURE-FAIR implementers. [https://bit.ly/2OWWryg](https://bit.ly/2OWWryg)
10 Things for Curating Reproducible and FAIR Research

- Standards-based guidelines for CURE-FAIR best practices in publishing and archiving computationally reproducible studies.
- Focus on social science research that relies on quantitative data to produce results.
- Intended audience:
  - Data curators and information professionals who are charged with verifying that a computation can be executed and can reproduce prespecified results.
  - Researchers, publishers, editors, reviewers, and others who have a stake in creating, using, sharing, publishing, or preserving reproducible research.
10 Things for Curating Reproducible and FAIR Research

- **Introduction**
  - Defines common terms
  - Identifies activities relevant to curation and computational reproducibility
  - Describes purpose of the 10 Things for Curating and FAIR Research

- **Table of Contents**
  - Explains the format of the document
    - Each 'Thing' begins with a section describing the issue followed by three actionable sections: 'Get started', 'Learn more', and 'Go deeper'
    - Highlights the main question each 'Thing' addresses
10 Things for Curating Reproducible and FAIR Research

- Questions addressed by Things
  - Does the research compendium contain everything needed to reproduce a predefined outcome in an organized and parsimonious way?
  - Is descriptive information about the research compendium and its components available and easy to understand?
  - Is information about how the research compendium and its components can be used available and easy to understand?
  - Is information about the research and its components embedded in code?
  - Is there a plan for reviewing the research compendium for FAIR and computational reproducibility standards over time?
## 10 Things for Curating Reproducible and FAIR Research

<table>
<thead>
<tr>
<th>1. Completeness</th>
<th>The research compendium contains all of the objects needed to reproduce a predefined outcome.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Organization</td>
<td>It is easy to understand and keep track of the various objects in the research compendium.</td>
</tr>
<tr>
<td>3. Economy</td>
<td>Fewer objects in the compendium mean fewer things that can break and less ongoing maintenance.</td>
</tr>
<tr>
<td>4. Transparency</td>
<td>The full context necessary to understand the research process is available.</td>
</tr>
<tr>
<td>5. Documentation</td>
<td>The process and reasoning required to reproduce a scientific claim are readily available and understandable.</td>
</tr>
<tr>
<td>6. Access</td>
<td>It is clear who can use what, how, and under what conditions, with &quot;open&quot; being preferred.</td>
</tr>
<tr>
<td>7. Provenance</td>
<td>The origin of the components of the compendium and how each has changed over time is evident.</td>
</tr>
<tr>
<td>8. Metadata</td>
<td>Information about the compendium and its components is embedded in a standardized schematic code.</td>
</tr>
<tr>
<td>9. Automation</td>
<td>As much as possible, the computational workflow is script-based to allow re-execution with minimal actions.</td>
</tr>
<tr>
<td>10. Review</td>
<td>A series of managed activities are in place to ensure continued access to and functionality of the compendium.</td>
</tr>
</tbody>
</table>
CURE-FAIR Working Group Recommendation

10 Things for Curating Reproducible and FAIR Research

The Challenge:
There is a lack of standardized approaches for curating materials and achieving a FAIR and reproducible scholarly record.

Produced by: CURE-FAIR WG
https://www.rd-alliance.org/groups/cure-fair-wg

What is the solution?
Developing a resource with standards-based guidelines for curating research corpora. The “10 Things for Curating Reproducible and FAIR Research” describe key practices for curating reproducible and FAIR research (CURE-FAIR), focusing primarily on research corpora produced by quantitative data-driven social sciences. Following the guidelines will improve the prospects for a reproducible scholarly record.

What is the impact?
Researchers, publishers, editors, reviewers, and others who have an interest in creating, using, sharing, publishing, or preserving reproducible research will be able to use the “10 CURE-FAIR Things” as guiding principles for FAIR and reproducible research. Data curators and information professionals who are charged with publication and archiving of research materials will be able to follow the guidelines and ensure these materials are FAIR and computationally reproducible.
10 Things for Curating Reproducible and FAIR Research

Website: https://curating4reproducibility.org/10things/
10 Things for Curating Reproducible and FAIR Research

- **Modular**
  - Select which Things are relevant to your needs
  - Select which level of activity matches your experience

- **Customizable**
  - Social Sciences
  - Potential for broader use in life sciences, long tail small scale data analysis

Data archives, institutional repositories, scientific publishers, data curators and archivists, and researchers willing to adopt the guidelines will benefit from a standardized approach to CURE-FAIR and what we hope will spur the development of widely-applicable solutions.
Odum Institute, University of North Carolina at Chapel Hill

- Adoption of the full CURE-FAIR recommendation as part of the corpus of standards and best practices that inform current data curation workflows
- "10 Things" to be used in training programs as a framework for understanding critical aspects of data curation that support reproducibility

Cornell Center for Social Sciences, Cornell University

- As an implementer of CURE-FAIR practices, recommendations have already been applied to curation activities
- The "10 Things" articulates the principles that inform the Results Reproduction (R-squared) service
Endorsement and Adopters

Cornell Center for Social Sciences, Cornell University

- Results Reproduction Service
  - Level of adoption: Production
  - Continue to use recommendation: Yes
  - Modified the recommendations: No
  - Time/effort spent on adoption: Minimal
  - Overall experience: Very good!
  - Do it again? Certainly! Over and over

- Data and Reproduction Archive  https://archive.ciser.cornell.edu/
  - Type "R2" in the search box

- Sample study highlighting the 10 Things:
Using 10 CURE-FAIR Things

- **Training researchers and data support professionals**
  10 CURE-FAIR Things provides an instructional outline for teaching essential concepts and skills required for curating data for reproducibility.

- **Developing and evaluating data curation workflows**
  10 CURE-FAIR Things can inform the development and evaluation of CURE-FAIR workflows by defining the goals of curation activities.

- **Articulating the purpose of data management policies and their implementation**
  10 CURE-FAIR things describes standards-based guidelines that inform requirements of data policies that promote reproducibility
  - For example: AJPS verification policy and verification workflow
Future of 10 CURE-FAIR Things

Recommendation (DOI: 10.15497/RDA00074)
- Maintenance mode
- Odum Institute

Supporting outputs
- Challenges
  DOI: 10.15497/RDA00063
- Bibliography
  https://www.zotero.org/groups/2868459/rda_cure-fair_subgroup_1/library
- Practitioner survey (ongoing)
Future of 10 CURE-FAIR Things

- 10 Things for Curating Reproducible and FAIR Research: https://curating4reproducibility.org/10things/
- GitHub: https://bit.ly/3n5miS5
- How to contribute:
  - Open an issue to suggest edits
  - Make edits and submit a pull request
  - Fork the GitHub repository to create a new discipline-specific version
Future of 10 CURE-FAIR Things

Opportunities

- License
  - CC-BY
- Length
  - Cutting is easier than writing
- Format
  - 3 different forms for now
- Detail
  - Can be updated
<table>
<thead>
<tr>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adil Hasan</td>
</tr>
<tr>
<td>Alain Marois</td>
</tr>
<tr>
<td>Alex May</td>
</tr>
<tr>
<td>Alicia Martinez-Garcia</td>
</tr>
<tr>
<td>Amy Nurnberger</td>
</tr>
<tr>
<td>Ana Lucic</td>
</tr>
<tr>
<td>Ana Van Gulick</td>
</tr>
<tr>
<td>Ana Trisovic</td>
</tr>
<tr>
<td>André Pacheco</td>
</tr>
<tr>
<td>Andrea Ogier</td>
</tr>
<tr>
<td>Andrew Davison</td>
</tr>
<tr>
<td>Andy Gotz</td>
</tr>
<tr>
<td>Ann Borda</td>
</tr>
<tr>
<td>Annajiat Alim Rasel</td>
</tr>
<tr>
<td>Anthony Juehne</td>
</tr>
<tr>
<td>Björn Brötz</td>
</tr>
<tr>
<td>Brian Alper</td>
</tr>
<tr>
<td>Brian Westra</td>
</tr>
<tr>
<td>Carmen Reverté</td>
</tr>
<tr>
<td>Celia Alvarez-Romero</td>
</tr>
<tr>
<td>Chenzhou Cui</td>
</tr>
<tr>
<td>Chinmaya Kumar Dehury</td>
</tr>
<tr>
<td>Daniel Nüst</td>
</tr>
<tr>
<td>Daniel S. Katz</td>
</tr>
<tr>
<td>Dieuwertje Bloemen</td>
</tr>
<tr>
<td>Elli Papadopoulou</td>
</tr>
<tr>
<td>Erin Clary</td>
</tr>
<tr>
<td>Fernando Aguilar</td>
</tr>
<tr>
<td>Florio Arguillas</td>
</tr>
<tr>
<td>Frankie Stevens</td>
</tr>
<tr>
<td>Frederik Geth</td>
</tr>
<tr>
<td>Gerard Coen</td>
</tr>
<tr>
<td>Halle Burns</td>
</tr>
<tr>
<td>Hervé L'Hours</td>
</tr>
<tr>
<td>Hilary Hanahoe</td>
</tr>
<tr>
<td>Hilary Szu Yin Shiue</td>
</tr>
<tr>
<td>Ingmars Kreismanis</td>
</tr>
<tr>
<td>Irina Hope</td>
</tr>
<tr>
<td>Jake Carlson</td>
</tr>
<tr>
<td>James Wilmot</td>
</tr>
<tr>
<td>James Hester</td>
</tr>
<tr>
<td>Jeffrey Grethe</td>
</tr>
<tr>
<td>Jennifer Harrow</td>
</tr>
<tr>
<td>Jennifer Huck</td>
</tr>
<tr>
<td>Joanne Fitzpatrick</td>
</tr>
<tr>
<td>Jose Benito Gonzalez Lopez</td>
</tr>
<tr>
<td>Josh Moore</td>
</tr>
<tr>
<td>Karsten Peters-von Gehlen</td>
</tr>
<tr>
<td>Katherine Mika</td>
</tr>
<tr>
<td>Keith Russell</td>
</tr>
<tr>
<td>Lars Vilhuber</td>
</tr>
<tr>
<td>Laura Molloy</td>
</tr>
<tr>
<td>Laurent Heirendt</td>
</tr>
<tr>
<td>Lesley Wyborn</td>
</tr>
<tr>
<td>Limor Peer</td>
</tr>
<tr>
<td>Lisa Johnston</td>
</tr>
</tbody>
</table>
Louise Gillis
Madiareni Sulaiman
Maja Dolinar
Mandy Gooch
Maria Shatz
Markus Englund
Martin Thomas Horsch
Martina Stockhause
Mary Uhlmansiek
Matthias Liffers
Melissa Cuthill
Melissa Cuthill
Michelle Barker
Mikala Narlock
Morane Gruenpeter
Nadica Miljković
Naeem Muhammad
Naomi Peck
Paula Andrea Martinez
Peter McQuilton
Philipp Conzett
Qian Zhang
Rebecca Grant
RITA MENESES
Robert Downs
Roberto Di Cosmo
Rochelle Palmer
Romain DAVID
Rossella Aversa
Sa-kwang Song
Sandor Brockhauser
Sarah Callaghan
Seraphim Alvanides
Sergio Serra
Shahira Khair
Shanmugasundaram
Venkataraman
Shelley Stall
Sherry Lake
Sinziana Paltineanu
Sirko Schindler
Stefanie Kethers
Stephanie Hagstrom
Stephanie Holmgren
Stian Soiland-Reyes
Susan Ivey
Susana Barbosa
Sylvie Cocaud
Thu-Mai Christian
Tim Dennis
Tom Honeyman
Tomasz Miksa
Tristan Matthews
Udayanto Dwi Atmojo
Vaidas Morkevicius
Viviana Letizia
Wanda Marsolek
Wolmar Nyberg
Åkerström
Yaqin Yuan
Yasuyuki Minamiyama
Yvan Le Bras
Zhifang Tu
The WG - Subgroup Members & Contributors

- Erin Clary, Christopher Erdmann, Ana Van Gulick, Daniel S. Katz, Katherine E. Koziar, Wanda Marsolek, Peter McQuilton, Qian Zhang

- Nadica Miljković, Karsten Peters-von Gehlen
THANK YOU!

CURE-FAIR WG webinar
Thursday, 22 September 2022
Florio Arguillas, Thu-Mai Christian, Mandy Gooch, Tom Honeyman, Limor Peer