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 & Research14% Public Administration11% Enterprise & Industry6% Other

# **RDA** Groups

33 Working Groups 53 Interest Groups

1 Community of Practice

rd-alliance.org

Ð



US

@resdatall | @RJA

BY SA

# 30 National Groups

Asia	Czech Republic	Italy	Slovenia			
Australia	Denmark	Lithuania	So Eastern Europe			
Austria	Estonia	Hungary	Spain			
Brazil	Europe	Ireland	Sweden			
Bulgaria	Finland	Netherlands	Region of the Americas			
Canada	France	Norway	United States			
Costa Rica	Germany	Portugal	United Kingdom			
Croatia	Greece	Romania	ontou ranguoni			
4 Contributing Regions						
Australia EU France United States (www.rd-alliance.org/groups/rda-us)						
www.rd-alliance.org/groups/national-groups						
rd-alliance.org						

@resdatall | @RLA\_US

BY SA

 $\odot$ 

B





rd-alliance.org @resdatall | @RDA





#### **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.



### Recording

Please note, this event is being recorded,

and will be available on the RDA website.



Recording 🗲

#### **Questions and Answers**

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







rd-alliance.ord



RESEARCH DATA ALLIANCE

Introducing 10 Things for Curating Reproducible and FAIR Research Webinar



**Limor Peer** 



Mandy Gooch



Tom Honeyman



**Hosted by RDA-US** 







@resdatall | @REA US



rd-alliance.org

@resdatall

UNITED STATES

Hosted by RDA-US

# Thanks to....

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

@R

# www.rd-alliance.org/groups/rda-us



Hosted by RDA-US



#### **Upcoming Webinars in this 2022 Series**

• Research Metadata Schemas WG - October 25 at 20:00 UTC

rd-alliance.ord

• DMP Common Standards WG - November 22 at 17:00 UTC

# www.rd-alliance.org/groups/rda-us

@resdatall | @R





# 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



B





# Research Data Alliance 10th Anniversary Plenary meeting

#### Gothenburg, Sweden, March 21-23, 2023





10



UNITED STATES

Hosted by RDA-US

#### **RDA Global**

- Web www.rd-alliance.org
- Email enquiries@rd-alliance.org
- **Twitter @resdatall**

**RDA US** 

Email – rdaus@rda-foundation.org Twitter - @RDA\_US Web - www.rd-alliance.org/groups/rda-

- LinkedIn www.linkedin.com/in/ResearchDataAllia H&e
- **Slideshare www.slideshare.net/ResearchDataAlliance**

6

rd-alliance.org

# **Thank You For Joining Us!**

@resdatall | @R





## **CURE-FAIR WG webinar**

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

Thursday, 22 September 2022

research data sharing without barriers rd-alliance.org



Time	Agenda	
3 minutes	Introduction of RDA and RDA-US/Stephanie Hagstrom	
2 minutes	Introduction of Webinar and Presenters/Tom Honeyman	
40 minutes	es CURE-FAIR WG presentation	
	<ul> <li>About the CURE-FAIR WG: Objectives, case statement, supporting outputs/Limor Peer</li> </ul>	
	<ul> <li>10 Things for Curating Reproducible and FAIR Research/Mandy Gooch</li> </ul>	
	Adopter stories/Florio Arguillas	
	Using 10 CURE-FAIR Things/Thu-Mai Christian	
	Future of 10 CURE-FAIR Things/Tom Honeyman	
10 minutes	Q&A: Moderator/Tom Honeyman	
2 minutes	Closing/Stephanie Hagstrom	



### 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



- 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.



## **Community Impact**

- 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: <u>10.15497/RDA00063</u>
- 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. <u>https://www.zotero.org/groups/2868459/rda\_cure-fair\_subgroup\_1/library</u>
- 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



- 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.



- 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



- 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?



### **CURE-FAIR WG Final Recommendation**

Get Started

× × ×

Learn More

××

Go Deeper

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

#### **CURE-FAIR Working Group Recommendation**

Arguillas, F., Christian, T., Gooch, A., Honeyman, T., Peer, L., & CURE-FAIR WG. (2022). 10 Things for Curating Reproducible and FAIR Research (1.1). <u>https://doi.org/10.15497/RDA00074</u>







#### **10 Things for Curating Reproducible and and FAIR Research**

Website: https://curating4reproducibility.org/10things/

#### Github: https://bit.ly/3n5miS5

CURATING /~ About V News CURE Training 10 Things Resources ~ 5 G 10 Things for Curating Reproducible and FAIR Research CURE Training 10 Things Resources ~ Computational reproducibility requires a village curators and information professionals who are 10 THINGS Thing 5: Documentation Overview computation can be executed and can reprodu-Introduction Related FAIR Principles: F2, R1, R1.2, R1.3 is for researchers, publishers, editors, reviewer Thing 1: Completeness Related FAIRARS Principles: R1, R1.2, R3 creating, using, sharing, publishing, or preservi Thing 2: Organization Thing S: Economy An important strategy for demonstrating research transparency is to create Thing 4: Transparency comprehensive documentation that captures important details of the research process. Thing 5: Documentation so that others are able to follow it to reproduce published research results. Failing to Thing 6: Access record essential information about the research compendium-descriptions of Thing 7: Provenance individual compandium files, technical requirements for using compandium files, the Overview → Thing 8: Metadata origin of materials used in secondary analyses, dataset variable and value code Thing 9: Automation definitions, analytical steps for generating expected results--relegates the research Thing 10: Review process to a "black box," making it difficult or even impossible to reproduce results. Introduction  $\rightarrow$ Robust documentation is the outcome of an intentional practice that keeps in mind the information needs of the compendium re-user so that the research compendium and the materials contained within it are independently understandable for informed reuse. Thing 1: Completeness → > Get Started > Learn More Go Deeper research data sharing without barriers There are several tools available to assist with creating documentation that is complete, accurate, and understandable. The resources below include software rd-alliance.org packages, tutorials, and other tools to go one step further toward research

transparency and reproducibility

\$7 G

- 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



#### **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 <u>https://archive.ciser.cornell.edu/</u>
  - Type "R2" in the search box
- Sample study highlighting the 10 Things: Meemken, et al. Reproduction Materials for: Smallholder Farmers and Contract Farming in Developing Countries. <u>https://doi.org/10.6077/190x-1677</u>



## **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: <u>10.15497/RDA00063</u>
- Bibliography <u>https://www.zotero.org/groups/2868459/rda\_cure-</u> <u>fair\_subgroup\_1/library</u>
- Practitioner survey (ongoing) <u>https://bit.ly/2OWWryg</u>



## **Future of 10 CURE-FAIR Things**

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

Search or jump to 7 Pull reques	sts Issues Marketplace Explore	+• 🐠
Curating4reproducibility / curating4reprodu	cibility.github.io 🛇 Sponsor 🛇 Edit Pins + 💿 Watch 1 + 🦞 Fork 0 🏠 Sta	r 0 -
Scode ⊙ Issues \$1 Pull requests 5 ⊙ Action	ns 🖽 Projects 🖽 Wiki 💿 Security 🗠 Insights 🕸 Settings	
ا در	tent / en / 10things / Go to file Add file	
Ichristian update content	√ a4ba549 13 days ago 🕻	History
🗅 _index.md	update layouts 14	days ago
🖺 intro.md	update content 13	days ago
🗅 overview.md	lint 13	days ago
🗅 thing01.md	lint 13	days ago
🗅 thing02.md	lint 13	days ago
🗅 thing03.md	fix tings meta 13	days ago
🗅 thing04.md	fix tings meta 13	days ago
🗅 thing05.md	fix tings meta 13	days ago
🗅 thing06.md	fix tings meta 13	days ago
🗅 thing07.md	lint 13	days ago
🗅 thing08.md	lint 13	days ago
🗅 thing09.md	update links 13	days ago
🗅 thing10.md	update links 13	days ago
© © 2022 GitHub, Inc. Terms Privacy Security		



## **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



### **The WG - Group/List Members**

Adil Hasan Alain Marois Alex May Alicia Martinez-Garcia Amy Nurnberger Ana Lucic Ana Van Gulick Ana Trisovic André Pacheco Andrea Ogier Andrew Davison Andy Gotz Ann Borda Annajiat Alim Rasel Anthony Juehne Björn Brötz Brian Alper **Brian Westra** Carmen Reverté

Celia Alvarez-Romero Chenzhou Cui Chinmaya Kumar Dehury Daniel Nüst Daniel S. Katz Dieuwertje Bloemen Elli Papadopoulou Erin Clary Fernando Aquilar Florio Arguillas Frankie Stevens Frederik Geth Gerard Coen Halle Burns Hervé L'Hours **Hilary Hanahoe** Hilary Szu Yin Shiue **Ingmars Kreismanis** Irina Hope

Jake Carlson James Wilmot James Hester Jeffrey Grethe Jennifer Harrow Jennifer Huck Joanne Fitzpatrick Jose Benito Gonzalez Lopez Josh Moore Karsten Peters-von Gehlen Katherine Mika Keith Russell Lars Vilhuber Laura Molloy Laurent Heirendt Lesley Wyborn I imor Peer Lisa Johnston



### **The WG - Group/List Members**

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-Reves 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<sup>26</sup>

- 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







27



# **THANK YOU!**

## **CURE-FAIR WG webinar**

#### Thursday, 22 September 2022

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



