

Navigating Data Sharing in International Research Collaborations

Hosted by RDA-US





Ingvill Constanze Mochmann

GESIS-Leibniz Institute for the Social Sciences **CBS International Business School** University of Oslo Germany and Norway



Kathleen Shearer **Confederation of Open** Access Repositories Canada

RESEARCH DATA ALLIANCE

UNITED STATES RD



Alison Specht The University of Oueensland Australia

Register Here: https://bit.ly/iN2N_Sept

September 16th **#RDAwebinar**











Membership

12,000+ members from 145 countries

Groups

97 Working and Interest Groups 30 National Groups 4 Regional Groups



rd-alliance.org

Øresdatall | @RDA_US





Recording

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

Navigating Data Sharing in International Research Collaborations

Recording

2021

Register Here

Global data communities are primed for considerable progress in accelerating discovery and advancing research through open science. However, navigating data sharing in the context of international collaborations can be extremely complex and very challenging. The International

Network-of-Networks (iN2N) Data Expert Workshop¹ was convened to address the social and technical challenges in making data shareable and interoperable internationally. The workshop brought together a multidisciplinary group of experts to identify the biggest challenges related to

Questions and Answers

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











Navigating Data Sharing in International Research Collaborations





Hosted by RDA-US











Upcoming Webinars Hosted by RDA-US

September 23, 2021 17:00 UTC

RDA Group Output Webinar Series

FAIR principles for research software (FAIR4RS WG)

October 7, 2021 20:00 UTC

RDA and ESIP Physical Samples Webinar Series

Supporting interdisciplinary sample data discovery, integration, and reuse October 21, 2021 15:00 UTC

> RDA Group Output Webinar Series

Getting attention for research project outputs and bringing others on board



@resdatall | @RDA_US



RDA RDA Plenary Meetings

RDA Plenaries are held every six months

18th Research Data Alliance Virtual Plenary Meeting

3 - 18 November 2021









(RDA) Preparation for RDA Plenary 18 Meeting

- What to Expect when Attending an RDA Plenary 28 Sep 2021 16:00 to 16:30 UTC
- Navigating All Things RDA Website, Plenaries and more 26 Oct 2021 16:00 to 16:30 UTC

www.rd-alliance.org/plenaries/rda-18th-plenary-meeting-virtual







RDA Plenary 18 Meeting

- **RDA Global**
- Email enquiries@rd-alliance.org
- Web www.rd-alliance.org
- Twitter @resdatall
- LinkedIn www.linkedin.com/in/ResearchDataAlliance
- **Slideshare www.slideshare.net/ResearchDataAlliance**

RDA US Email – rdaus@rda-foundation.org Twitter - @RDA_US

Thank You For Joining Us!











Social Challenges of Open Data Sharing and Perceptions of Risk

Robert R. Downs

rdowns@ciesin.columbia.edu

NASA Socioeconomic Data and Applications Center (SEDAC) Center for International Earth Science Information Network (CIESIN) Columbia Climate School, Columbia University

Navigating Data Sharing in International Research Collaborations Virtual Webinar Hosted by RDA-US September 16, 2021

Copyright 2021. The Trustees of Columbia University in the City of New York.



Beneficiaries of Open Data Sharing



Benefits of Ethical and Interoperable Open Data Sharing and Use

- Unrestricted access and use
 - Freely available for any type of use or user
 - Rights to use, modify and redistribute data
 - Enabling integration across catalogs, products, and services
- Open science
 - Fostering scientific progress
 - Contributes to transparency of the data and the results
 - Replicability and reproducibility
- Enabling other kinds of uses
 - Planning and policy making
 - Teaching and learning
- Contagion
 - Beneficiaries of open data sharing may also share their data
 - Recipients of open data may want to recognize those who share data



<u>()</u>

Social Challenges for Ethical and Interoperable Open Data Sharing and Use

- Changing behavior
 - Entrenched practices of not sharing data
 - Traditional science training did not emphasize open data sharing
- Potential risks for open data sharing
 - Inappropriate, misuse or misattribution of data
 - Possible missed opportunity of creators' competitive advantage
- Responsibilities of using open data
 - Respecting privacy and confidentiality
 - Not divulging sensitive data or personally identifiable information (PII)
- Effort to make data usable by others
 - Need to deposit data in an open repository
 - Need to clean, describe and document data, including open license and data quality
- New opportunities and challenges for sharing benefits from data sharing
 - Providing attribution for data producers
 - Rewarding (and not rewarding) data producers for sharing data



• 4

International Context for Social Challenges to Open Data Sharing

- Cultural differences
 - Data sharing practices and traditions
- Jurisdictional licensing
 - National laws
- Policies
 - National, local, and institutional
- Data description
 - Enabling use across domains, languages, professions and levels of expertise
- Practice
 - Divergent definitions and requirements for open data
 - Defining open data as "freely available data for anyone to use for any purpose without restrictions" is not universally accepted





Types of Perceived Risks of Sharing Data in an Open Manner

- Breach of confidentiality
 - Identification of personal, private, sensitive, or proprietary information
- Liability
 - Responsible for costs incurred from data use (or misuse)
- Missed opportunity
 - Others publishing research results before the data producers
- Misuse
 - Erroneous conceptualization, analysis, integration, and reporting
- Resource Commitment
 - Prepare, manage, preserve and persistently enable and support data use by others
- Misattribution
 - Absent, incomplete or incorrect reference to data
- Disparagement

6

• Criticism of data collection, processing, analysis, or interpretation







Adopting Principles to Reduce Perceived Data Sharing Risks and Increase Data Sharing Rewards



Group on Earth Observations

Data Sharing Principles Data Management Principles

Open Data Sharing to Reduce Risks and Foster Benefits

• Assign an open license to data



- Some open licenses offer liability protection and require attribution
- Clearly communicate rights and responsibilities to use open data
 - Data providers enable free use if users cite the data they use
- Encourage users to correctly cite the data used for publications
 - Data distributors offer recommended data citation on the data landing page



- Institutions reward data sharing as scientific contributions
 - Funding panels and promotion committees recognize data sharing

Data Citation Rewards Open Data Sharing

- Article citations are recognized metrics
 - Citations are counted for hiring and promotion
 - Departments and schools are evaluated with citations
 - Funders are evaluated with citations
- Multiple recipients when rewarding with data citations
 - Data producers and employers
 - Data users are recognized for properly citing data
 - Data repository and staff
 - Sponsors and program managers are increasing value of research funds
 - Journals and journal editors are recognized for their high standards





Open Data and Data Citations

- Need to improve the status of data citations
 - Increasing article authors' data citation practices can improve rewards
 - Correctly citing data improves capabilities to count data citations
 - If citing data use becomes common practice, data citations will be valued
- Data citations are earned
 - Preparing data for reuse improves the usability of open data
 - Sharing open and reusable data improves the efficiency of reuse
- Implications of repository choice
 - Some selective repositories conduct peer-review of data
 - Domain repositories serve particular communities of data users



Stakeholders Reduce Risks to Open Data Sharing

- Data Producers
 - Assign an unrestricted open license to data and document data quality
- Data Stewards
 - Clearly describe rights, conditions, limitations and appropriate uses
- Data Users
 - Correctly cite data used for a publication
- Sponsors
 - Recognize data sharing practices when funding research
- Employers
 - Give credit for sharing data and for data that have been cited
- Publishers and editors
 - Guide authors on data citation practices

Authors, Date, Title, Publisher, Persistent Identifier, Access Date



Gesis Leibniz Institute for the Social Sciences



"Navigating Data Sharing in International Research Collaborations" hosted by RDA-US, Webinar, 9/16 2021

Ingvill Constanze Mochmann GESIS-Leibniz Institute for the Social Sciences, Germany University of Oslo, Norway CBS International Business School, Germany



Jurisdictional, policy and cultural challenges of data sharing in international research collaborations

Placing the presentation in context

- Some experienced bottom-up controversies in research collaborations:
- What are research data?
- What is anonymisation?
- What is ethical?
- What is required to share data?
- How much effort and money should be put into making data sharable?
- Who wants to share?
- Who do the data belong to?
- Who is an author?

.









Countries have different laws and regulation with regard to for example

- Procedure for research project approval incl. data archiving, access & sharing of collected data
- Ethical standards & requirements
- Processes for violation of regulations

Furthermore, countries and disciplines have different traditions and cultures regarding sharing of data

 \rightarrow needs to be addressed and communicated from the start!





THANK YOU!

QUESTIONS?

Contact: ingvill.mochmann@gesis.org







Key elements to achieve FAIR outcomes in international multi-disciplinary groups

Alison Specht

Terrestrial Ecosystem Research Network, the University of Queensland, Australia













I would like to acknowledge the Traditional Owners and Custodians in all nations. I honour their profound connections to land, water, biodiversity and culture and pay my respects to their Elders past, present and emerging.



• groups of experts who meet face to face or remotely, over a sustained period of time to tackle topics of concern which each group member or their organisations have a desire to solve, or at least advance.



- groups of experts who meet face to face or remotely, over a sustained period of time to tackle topics of concern which each group member or their organisations have a desire to solve, or at least advance.
- members are usually chosen for their knowledge and expertise, and perhaps their ability to represent a vested interest.



- groups of experts who meet face to face or remotely, over a sustained period of time to tackle topics of concern which each group member or their organisations have a desire to solve, or at least advance.
- members are usually chosen for their knowledge and expertise, and perhaps their ability to represent a vested interest.
- in the environmental and earth sciences national boundaries are irrelevant to most topics, and often few group members will know one another personally.



- groups of experts who meet face to face or remotely, over a sustained period of time to tackle topics of concern which each group member or their organisations have a desire to solve, or at least advance.
- members are usually chosen for their knowledge and expertise, and perhaps their ability to represent a vested interest.
- in the environmental and earth sciences national boundaries are irrelevant to most topics, and often few group members will know one another personally.
- the groups are outcome-oriented. For researchers, publications, impact factors and citations are the goal. Same applies to data and code!

Interdisciplinary, international (working) groups





n









From Specht & Crowston, subm. Both p < 0.05

a secure (data) working environment



Principles

- Transparency
- Ethical behaviour
- Respect

How?

- Shared work platform
- Shared storage space
- Regular communication
- A system of versioning
- Accommodate time zones

S









First, focus on the researcher

First, focus on the researcher

F.A.I.R.

ORCID profile – ensure everyone has one

- Activating the automatic updates from ______
 Crossref (published papers) and DataCite (published datasets and other digital ______
 objects)
- + More info: http://bit.ly/ORCID_Trust



DDOMP (or DMP)

F.A.I.R.

Make one!

A data mentor* will help the team start their DDOMP and set up easy-to-follow regular reviews for dedicated team members to follow.

e.g. Stall et al. 2021: Version 2, <u>https://doi.org/10.5281/zenodo.4910115</u>

* every project should have one!

https://bfe-inf.github.io/toolkit/ddomp.html





F.A.I.R.

Using and producing data

- Check licences to use others' data and code, and choose the license for your own data (and code) (<u>https://chooser-beta.creativecommons.org/</u>)
- attribution requests are recorded, and provenance notes taken



F.A.I.B.

Using and producing data and code

- Check permissions to use others' data and code, and choose the license for your own (<u>https://chooserbeta.creativecommons.org/</u>)
- attribution requests are recorded, and provenance notes taken

Choose a home for your data and code

- a TRUSTed* repository
- In the interim, safe but shared storage sites for work-in-progress

* https://codata.org/rdm-glossary/trusted-digital-repository/





F.A.I.B.



Describe the data and code

- Use community-accepted, semantic vocabularies;
- Rich metadata. Make sure metadata is adequate for re-usability / reproducibility____
- Use a recognised dialect (e.g. EML, ISO)
- Include clear licensing statements



F.A.I.R.

publishing data and code

- Get a unique persistent identifier (a PID) for metadata and code
 - Obtain a **doi** for digital objects, papers, data, **purl** for code etc.





F.A.I.R.

publishing data and code

- Get a unique persistent identifier (a PID) for metadata and code
 - Obtain a **doi** for digital objects, papers, data, code, purl for vocabularies etc





This will link to the author's ORCID



Discovery, citations, re-use, funding, promotion!

PARSEC Team Resources

- Material development and temporary storage
- + Google Drive
 - Team communications and information decimation
- + Email, Slack

- < #
- □ Short-term Dataset storage (during project)
- + Open Science Framework (integrated with AWS)
- Data preservation (including derived products)
- + Environmental Data Initiative
- □ References
- + Zotero





- □ Software development
- + GitHub
- □ Software preservation
- + Zenodo (integrated with GitHub)
- zenodo

E S

PAR

- Training, workshop material preservation
- + Zenodo
- Credit and automated attribution
- + Up-to-date ORCID
- Acknowledgments

"This work is part of the Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas (PARSEC) project with funding provided by the Belmont Forum < through country Grant ###### >."









Navigating the Complexities of Data Sharing in International Research Collaborations: Results of the IN2N Data Sharing Workshop

Facilitator Kathleen Shearer, COAR Organizers Anthony Juehne, RDA US Leslie McIntosh, RDA US

International Network to Network (IN2N) Workshop

- Funded through NSF grant
- Convened in April 2019
- Aim: to address the socio and technical challenges in making data shareable and interoperable internationally

25 Participants representing different stakeholders, domains and regions

Workshop Participants: Oya Beyan, Louise Bezuidenhout, Marcela Alfaro Córdoba, Robert Downs, Rorie Edmunds, Faisal Fadlelmola, Mark Hahnel, Cynthia Hudson-Vitale, Sarah Jones, Mark Leggott, Steve MacFeely, Gerardo Machnicki, Devika Madalli, Ingvill Constanze Mochmann, Tom Orrell, Karen Payne, Ge Peng, Lindsay Poirier, Peter Pulsifer, Rob Quick, Erin Robinson, Hugh Shanahan, Maria Sorokina, Alison Specht

Peter Pulsifer, Research Scientist, University of Colorado Boulder



"The ubiquity of information and communications technology that has emerged in recent decades provides the human race with unprecedented opportunities for making data shareable and interoperable at a global scale, including between and among nation-states, non-state actors, and inclusive of global commons."

But... it's complicated

Two distinct scenarios of data sharing: 1. during the research project (sharing across research team) and 2. after the project is over (sharing with others)

Four category of challenges for international data sharing

- I. Technologies
- II. Standards
- III. Regulatory and Ethical RequirementsIV. Data Practices



Presenters



Kathleen Shearer Confederation of Open Access Repositories (COAR) Canada



Ingvill Constanze Mochmann

GESIS-Leibniz Institute for the Social Sciences CBS International Business School University of Oslo Germany and Norway



Robert R. Downs Center for International Earth Science Information Network (CIESIN) Columbia University



Alison Specht Honorary Associate Professor, TERN Ecosystem Research Analyst The University of Queensland Australia



Lindsey Poirier, Assistant Professor of Science and Technology Studies University of California Davis

"To advance international and interdisciplinary collaborations, we need to better understand the unique opportunities and challenges that researchers in many diverse contexts face at multiple scales. This will require devising and applying new frameworks for unpacking diverse data cultures and thinking through how the design of data infrastructure and the development of recommended data practices can be interwoven through them."



Guidance for Data Sharing in International Research Collaborations



#1

Build consensus and trust in the team on the need and benefits of data sharing



Document the relevant policy and legal requirements



Identify skills gaps and develop a plan to address them



Develop a data governance framework

Data governance framework:

- Specify the data management roles and responsibilities across the research team
- Describe the rights and restrictions related to data sharing based on the nature of the data and the communities with whom the team is working
- List the standards for data interoperability, discovery and re-use
- Identify the tools and infrastructure through which data will be shared across the team, and beyond the team
- Document the procedures for managing and sharing data during and after the project

Data governance framework:

The framework is presented at a very high level, and many projects that wish to implement the framework will need to be considered in greater detail in different areas, For example,

- What are the responsibilities of staff and investigators for data management? (Include time allocations, project management responsibilities, training requirements, and contributions of non-project staff. Name specific individuals where possible.)
- How will the PI(s) verify that the data generated are being managed according to this plan?
- At what point(s) in the project will this happen? Who is responsible for checking that the plan is being followed?
- Is there a formal process for transferring responsibility for the data should a PI or co-PI leave his or her institution?
- Who will have responsibility for decisions about the data once all the original personnel are no longer associated with the project? Is there a procedure in place for transferring responsibility once the original personnel are no longer available?
- Who will bear the cost associated with data preparation, management, and preservation?

Thanks!

Q&A