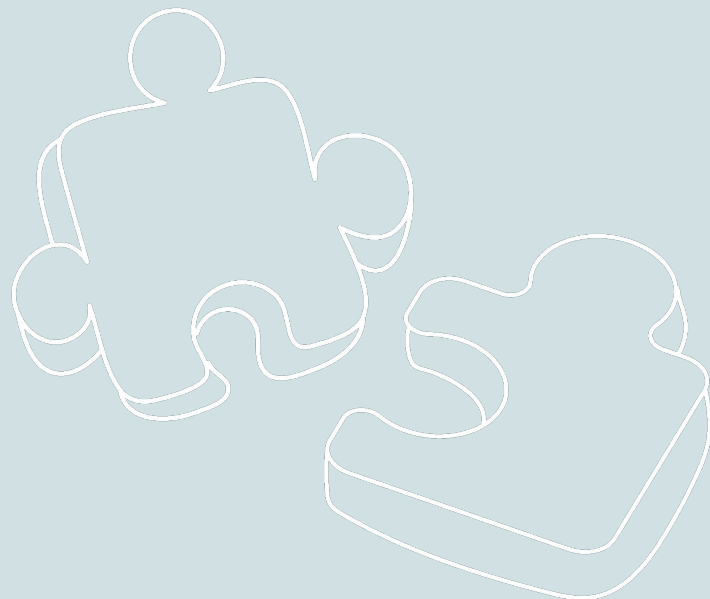


# Research Data Bootcamp: Data Publishing and CU Scholar



Andrew Johnson, Associate Professor, Director, Data and Scholarly Communication Services

Melissa Cantrell, Assistant Professor, Scholarly Communication Librarian



Center for Research Data & Digital Scholarship

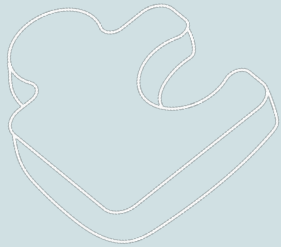
UNIVERSITY OF COLORADO **BOULDER**

# Agenda

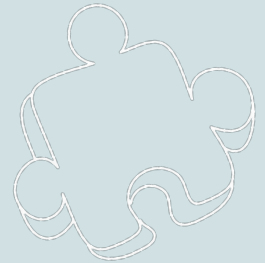
1. Introductions/ Who we are
2. What is data publishing?
3. Why publish data?
4. How to publish data
  - a. Intro to FAIR principles
  - b. Top considerations
  - c. CU Scholar demonstration
5. Questions and wrap-up



# 1. Introductions/Who We Are



## 2. What is data publishing?



# Working definition

- Making research data and metadata/documentation publicly available (or with appropriate access controls) via a formal web-based repository/database
- Preferably in adherence with [FAIR data principles](#) and/or other standards for data, metadata, and repository quality

# Related terms

- Data sharing
- Data curation
- Data archiving
- Data preservation

### 3. Why publish data?

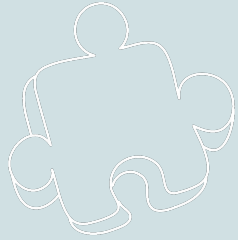


# Why publish data?

1. Scientific and public good
  - a. Advance scientific innovation
  - b. Address reproducibility
2. Journal/publisher requirements
  - a. “Data availability statements”
  - b. FAIR repositories with citations via persistent IDs
3. Funder requirements
  - a. Part of NSF data management plans since 2011



## 4. How to publish data



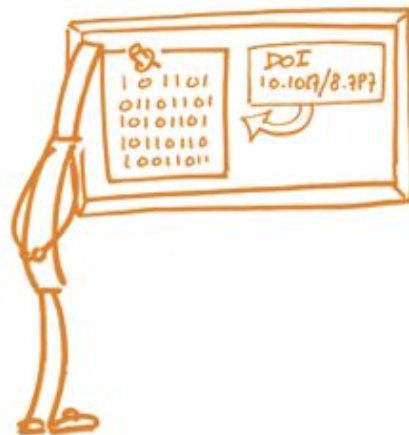
# Introduction to FAIR data principles ([Wilkinson et al., 2016](#))

## FAIR DATA PRINCIPLES

AH!



FINDABLE



ACCESIBLE

HOW DO YOU  
OPEN A .XZQ FILE?



INTEROPERABLE



REUSABLE

## Findable (F)

- Apply a globally unique and persistent identifier
- Describe your data in a data repository

### FINDABLE

Unique identifiers and metadata are used to allow data to be located quickly and efficiently



## Accessible (A)

- Consider what will be shared, and share via a open, free, and universally implementable protocol
- Metadata are valuable and accessible, even when the data are no longer available

### ACCESSIBLE

Data is open, free  
and universally  
available for  
research  
discovery efforts



# Interoperable (I)

- Use:
  - Open formats
  - Consistent vocabulary
  - Common metadata standards

## INTER-OPERABLE

A common programming language is used to allow use in a broad range of applications



## Reusable (R)

- Origin, context, history, and who to credit/cite are all crucial for data reusability
- Consider permitted used and apply the appropriate license

## REUSABLE

All data is clearly described and outlines associated data-use standards



# Top Considerations for You # 1

Have clear documentation and a data management plan from square one.

- Think ahead about repositories and requirements for a finished project
- Document throughout the process/project: how data was created/gathered/used/etc.
- R (Reusable) in FAIR is very hard to achieve just at the end of the project; important to think about from the beginning

# Top Considerations for You # 1

Before you start collecting data, think about:

- How much of your data will you share?
- How and where will you share your data?
- When will you share your data?
- With whom will you share your data?



## Top Considerations for You #2:

### Select a FAIR-aligned data repository

- CU Scholar
  - FAIR-aligned public access repository for CU Boulder-affiliated researchers (i.e., have a CU IdentiKey)
  - In process of obtaining CoreTrustSeal certification
  - Review and curation of all data sets
  - DataCite DOIs registered for all data sets
  - Public access to large data sets via Globus and PetaLibrary

# Top Considerations for You # 2:

## Select a FAIR-aligned data repository

- General repositories (e.g., [Dryad](#), [Dataverse](#), [Zenodo](#))
  - Open to anyone to deposit
  - Minimal review/curation of deposits
  - Typically provide DataCite DOIs and usage metrics
  - Size limits and/or additional fees for large data
- Domain repositories:
  - [Re3data](#) repository registry
  - Level of review/curation varies
  - Ability to deposit varies
  - May be recommended/required for certain data types by funders/publishers (e.g., [Springer-Nature's list](#))

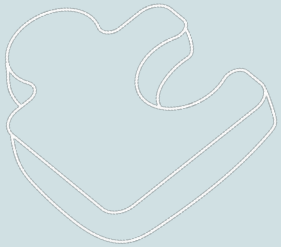
## Top Considerations for You # 3:



Consider copyright and licensing of your data set

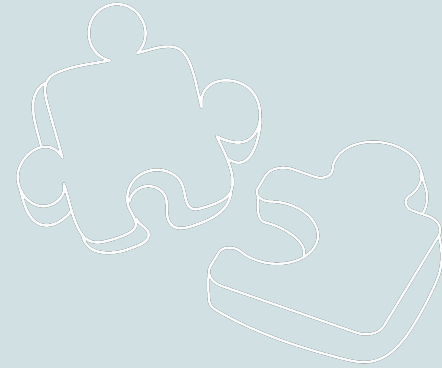
- Data itself (including facts and ideas) is not copyrightable
- The license that is selected facilitates sharing and reuse of the data set
  - [Creative Commons](#)
  - CC BY: Creative Commons Attributions License
  - CC 0: When an owner wishes to waive their copyright and/or database rights
  - Public Domain mark (PDM): It is used to mark works that are in the public domain, and for which there are no known copyright or database restrictions.

# CU Scholar demonstration



# Thank you!

Comments? Questions?



General Questions:  
[crdds@colorado.edu](mailto:crdds@colorado.edu)

Andrew Johnson:  
[andrew.m.johnson@colorado.edu](mailto:andrew.m.johnson@colorado.edu)

Melissa Cantrell  
[melissa.cantrell@colorado.edu](mailto:melissa.cantrell@colorado.edu)

CU Scholar Questions:  
[cuscholaradmin@colorado.edu](mailto:cuscholaradmin@colorado.edu)



Center for Research Data & Digital Scholarship

UNIVERSITY OF COLORADO **BOULDER**