Research Data Camp: Data Publishing and Repositories

Melissa Cantrell, Assistant Professor, Scholarly Communication Librarian

Matthew Murray, Teaching Assistant Professor, Data Librarian

August 20th, 2024

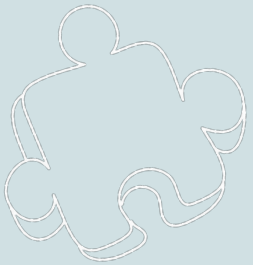
Center for Research Data & Digital Scholarship

UNIVERSITY OF COLORADO **BOULDER**

Agenda

1. What is data publishing?
2. Why publish data?
3. How to publish data
    a. Intro to FAIR principles
    b. Top considerations
    c. CU Scholar/Dryad examples
4. Questions and wrap-up

1. 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

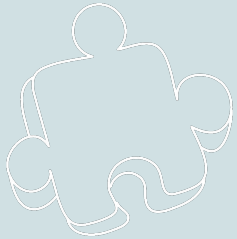
2. Why publish data?

Why publish data?

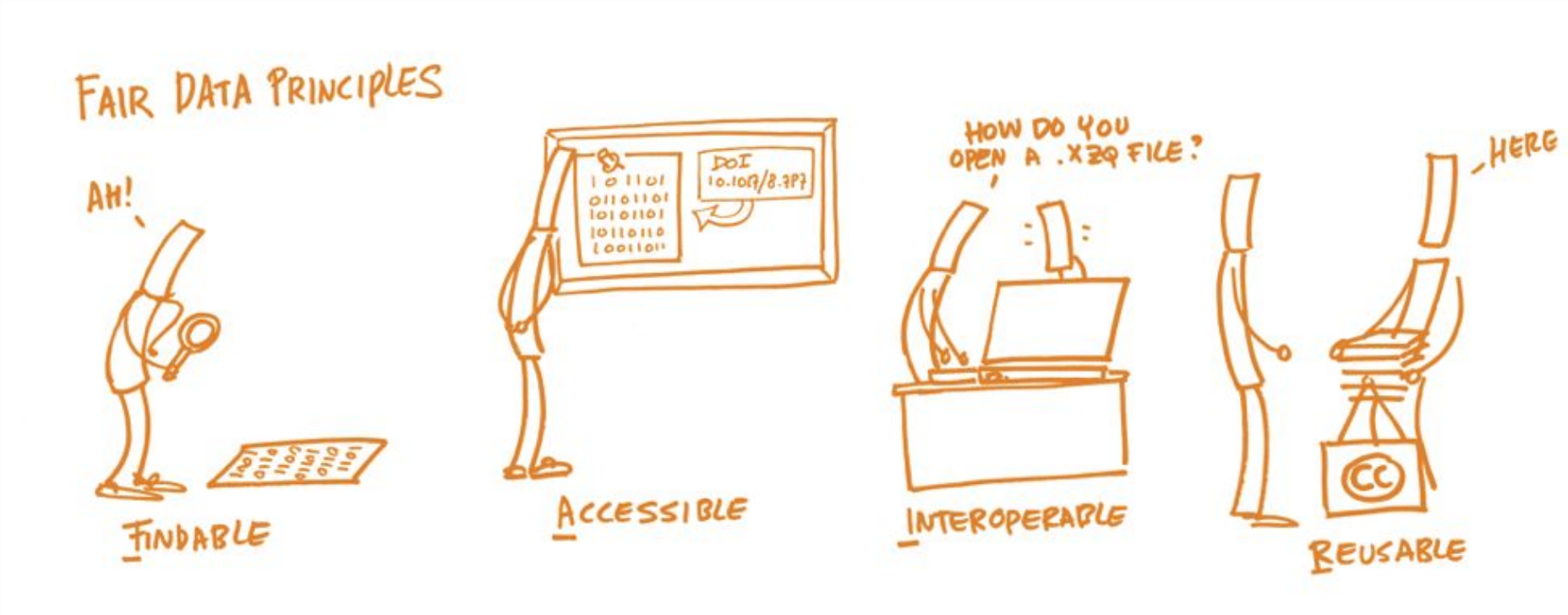
1. Scientific and public good
    a. Advance scientific innovation
    b. Address reproducibility
2. Journal/publisher requirements
    a. “Data availability statements”
    b. FAIR repositories with citations via persistent IDs
3. Funder requirements
    a. Part of NSF data management plans since 2011
    b. Part of NIH data management and sharing policy since 2023

Thinking Ahead

3. How to publish data



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

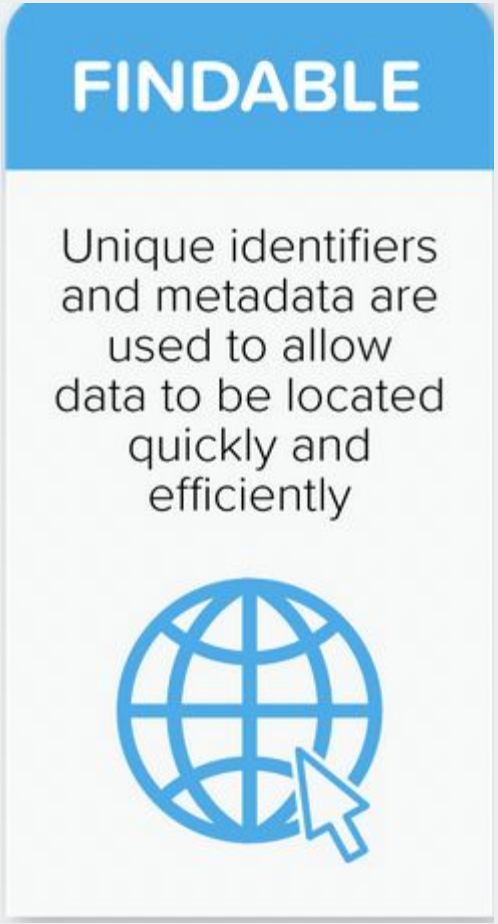


OpenAire: “How to Make Your Data FAIR”

Image by Patrick Hochstenbach, CC0 1.0

Findable (F)

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



https://www.go-fair.org/fair-principles/

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

https://www.go-fair.org/fair-principles/

Interoperable (I)

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

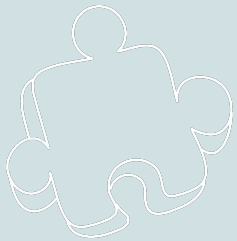
https://www.go-fair.org/fair-principles/

Reusable (R)

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

https://www.go-fair.org/fair-principles/

Top Considerations…**Before you start collecting data**



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/can you share?
- How and where will you share your data?
- When will you share your data?
- With whom will you share your data?

University of Colorado Boulder

CU Scholar

UNIVERSITY LIBRARIES

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 an IdentiKey)
  - Has 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
  - Free to deposit up to 500 GB per data set for CU Boulder affiliated researchers
  - Over 1600 data sets published in CU Scholar to date

Top Considerations for You # 2:

Select a FAIR-aligned data repository

- Dryad @ CSU
  - Dryad is a non-profit FAIR-aligned data repository
  - Data preserved in CoreTrustSeal-certified repository
  - Review and curation of all data sets
  - DataCite DOIs registered for all data sets
  - Free to deposit up to 300 GB per data set for CSU affiliated researchers
  - Requires ORCID for login
  - Over 400 CSU data sets published in Dryad

Top Considerations for You # 2:

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

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

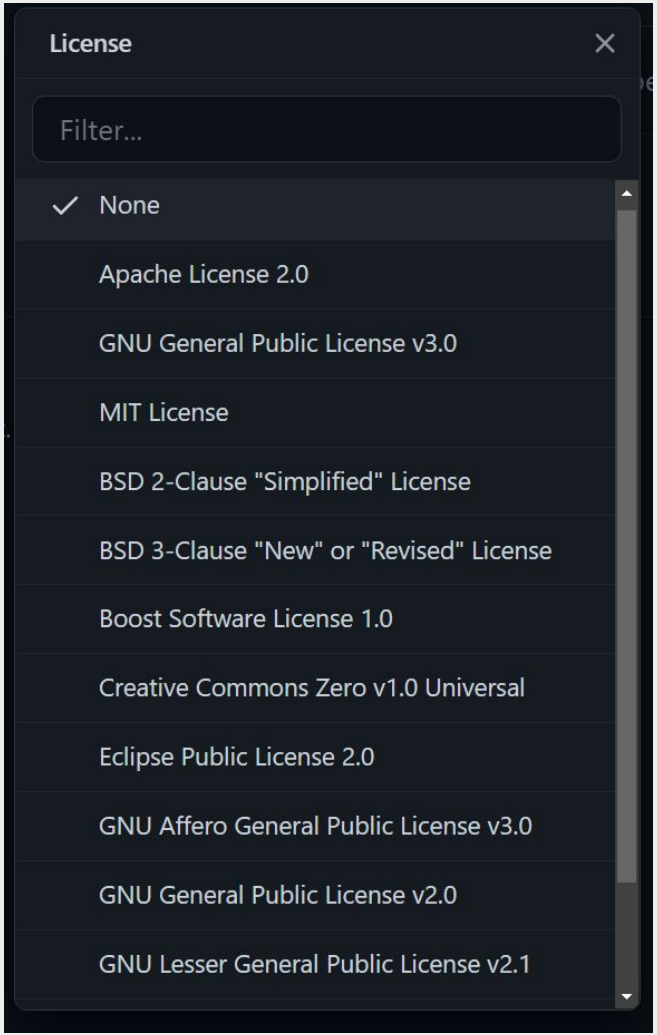
- The license that is selected facilitates sharing and reuse of the data set
  - Creative Commons
  - CC BY: Creative Commons Attributions License
  - CC 0: When an owner wishes to waive their copyright and/or database rights
  - Public Domain mark (PDM): It is used to mark works that are in the public domain, and for which there are no known copyright or database restrictions.

Top Considerations for You # 3:

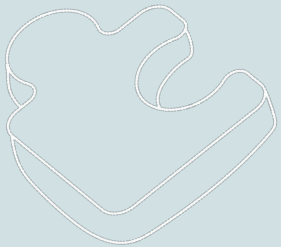
Consider copyright and licensing of associated software/code

- Many licenses available for software/code

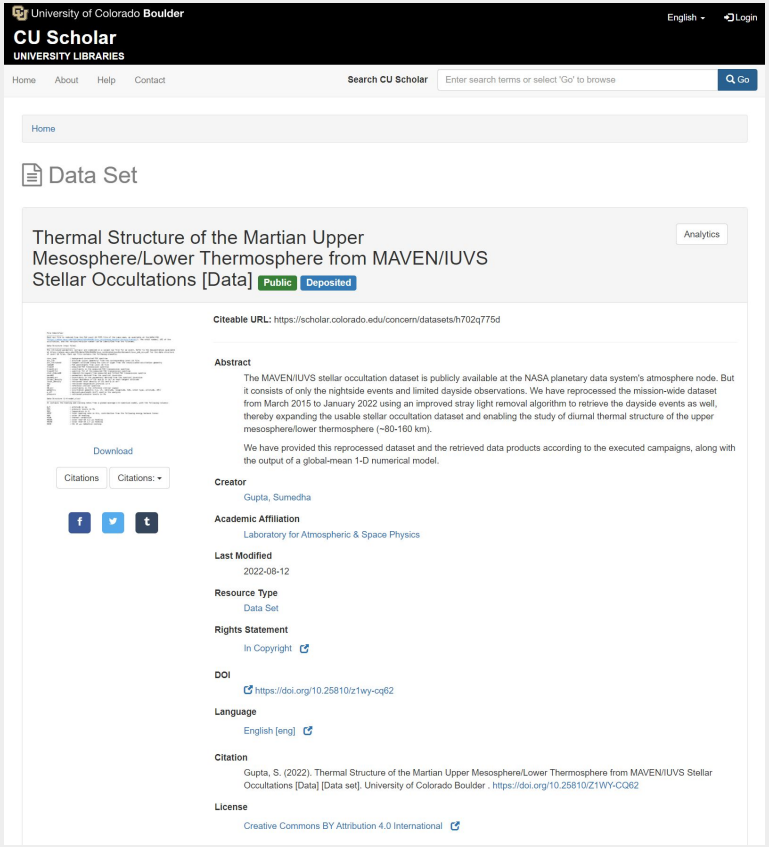
GitHub

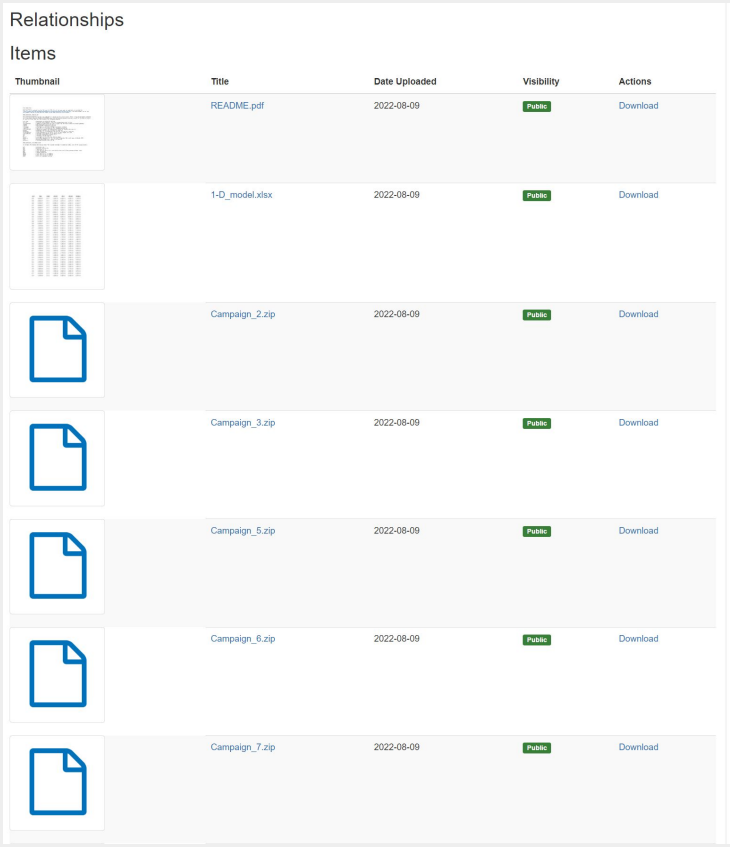


CU Scholar/Dryad Examples



Example Data Set in CU Scholar





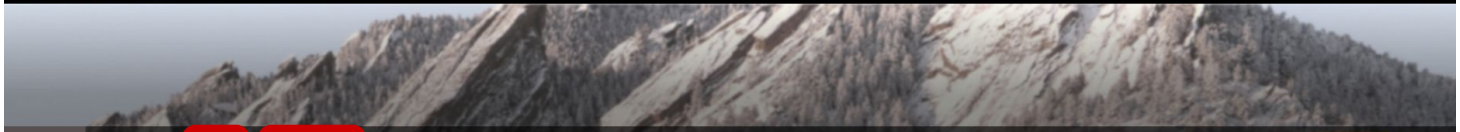
https://doi.org/10.25810/z1wy-cq62

University of Colorado **Boulder**

English | Matthew Murray

CU Scholar

UNIVERSITY LIBRARIES



HOME | ABOUT | HELP | CONTACT

Search CU Scholar | Enter search terms or select 'Go' to browse | Go | All

Share Your Work

Terms of Use

Featured Works | Recently Uploaded

Genetic and Environmental Etiologies of Reading Disabilities: Analysis of data from the Colorado Learning Disabilities Research Center

Creator: Astrom, Raven

Subject: Twins, Heritability, Longitidunal, Reading, Disability, Siblings

Resource Type: Dissertation

The 19 Percent: Disability and Actor Training in Higher Education

Creator: McNish, Deric

Subject: Universal Design for Learning, Performance, Disability, Voice, Actor Training

Resource Type: Dissertation

Exploring Web Simplification for People with Cognitive Disabilities

Creator: Hoehl, Jeffrey Arthur

Subject: human computer computing, web simplification, cognitive disabilities, accessibility, human computer interaction, web accessibility

Resource Type: Dissertation

Save order

Explore Collections | Featured Researcher

Works

Electromagnetics Laboratory/The MIMICAD Research Center

Series in Biology

University Libraries Fellows

Western States Government Information Virtual Conference

View all collections

University of Colorado **Boulder**

English | Matthew Murray

CU Scholar

UNIVERSITY LIBRARIES

HOME | ABOUT | HELP | CONTACT

Search CU Scholar | Enter search terms or select 'Go' to browse | Go | All

Share Your Work

Terms of Use

Featured Works | Recently Uploaded

Genetic and Environmental Etiologies of Reading Disabilities: Analysis of data from the Colorado Learning Disabilities Research Center

Creator: Astrom, Raven
Subject: Twins, Heritability, Longitidunal, Reading, Disability, Siblings
Resource Type: Dissertation

The 19 Percent: Disability and Actor Training in Higher Education

Creator: McNish, Deric
Subject: Universal Design for Learning, Performance, Disability, Voice, Actor Training
Resource Type: Dissertation

Exploring Web Simplification for People with Cognitive Disabilities

Creator: Hoehl, Jeffrey Arthur
Subject: human computer computing, web simplification, cognitive disabilities, accessibility, human computer interaction, web accessibility
Resource Type: Dissertation

Save order

Explore Collections | Featured Researcher

Works

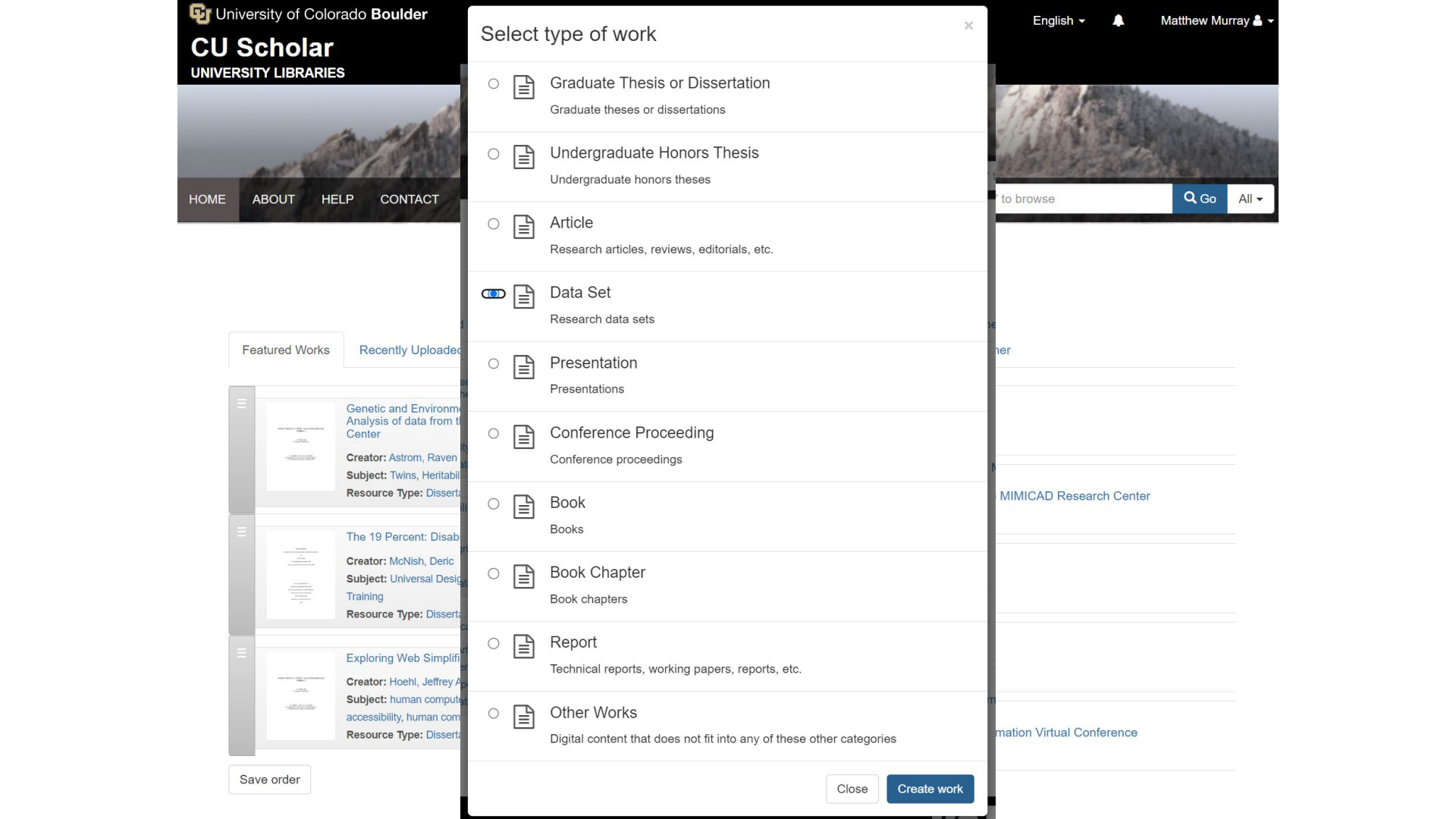
Electromagnetics Laboratory/The MIMICAD Research Center

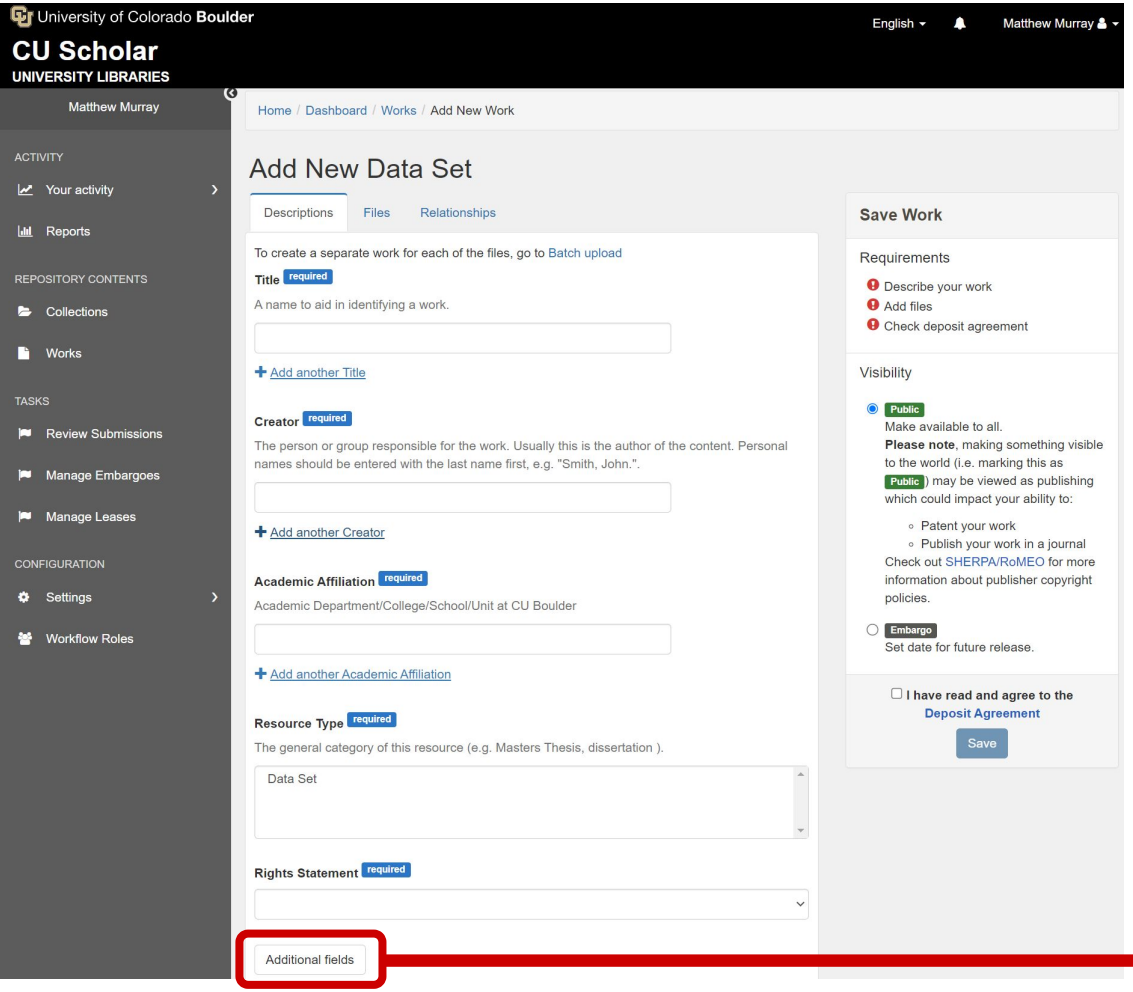
Series in Biology

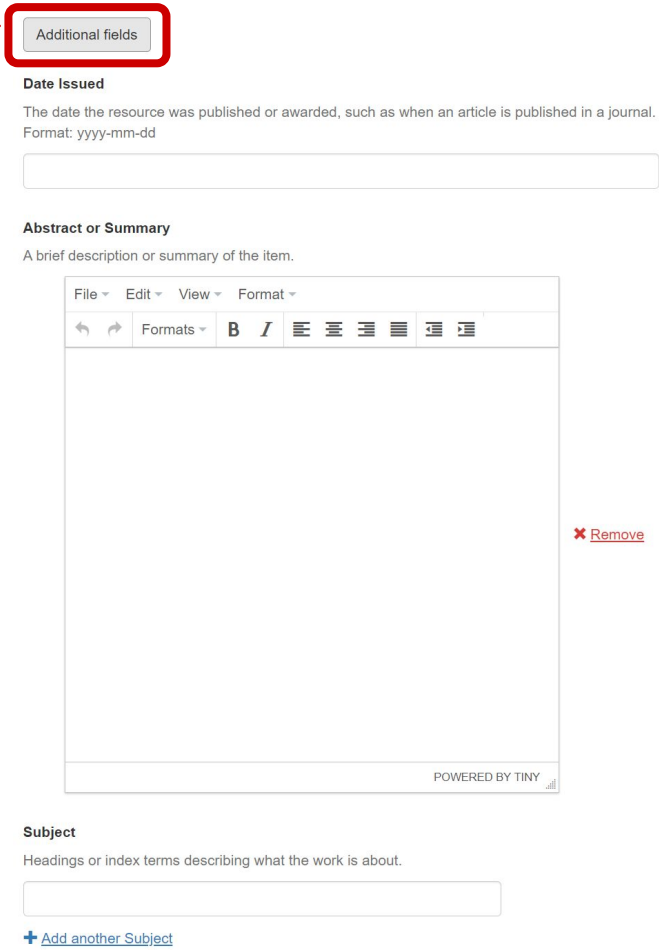
University Libraries Fellows

Western States Government Information Virtual Conference

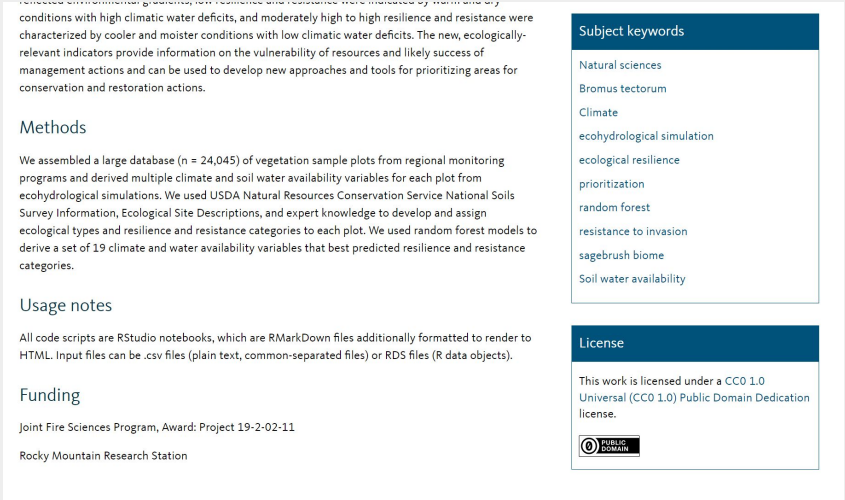
View all collections


University of Colorado Boulder
CU Scholar
UNIVERSITY LIBRARIES
English
Matthew Murray
HOME
ABOUT
HELP
CONTACT
to browse
Go
All
Select type of work
Graduate Thesis or Dissertation
Graduate theses or dissertations
Undergraduate Honors Thesis
Undergraduate honors theses
Article
Research articles, reviews, editorials, etc.
Data Set
Research data sets
Presentation
Presentations
Conference Proceeding
Conference proceedings
Book
Books
Book Chapter
Book chapters
Report
Technical reports, working papers, reports, etc.
Other Works
Digital content that does not fit into any of these other categories
Close
Create work
Featured Works
Recently Uploaded
Genetic and Environm
Analysis of data from t
Center
Creator: Astrom, Raven
Subject: Twins, Heritabil
Resource Type: Dissert
The 19 Percent: Disab
Creator: McNish, Deric
Subject: Universal Desig
Training
Resource Type: Dissert
Exploring Web Simplifi
Creator: Hoehl, Jeffrey A
Subject: human compute
accessibility, human com
Resource Type: Dissert
Save order
MIMICAD Research Center
mation Virtual Conference


University of Colorado Boulder
CU Scholar
UNIVERSITY LIBRARIES
English
Matthew Murray
Matthew Murray
ACTIVITY
Your activity
Reports
REPOSITORY CONTENTS
Collections
Works
TASKS
Review Submissions
Manage Embargoes
Manage Leases
CONFIGURATION
Settings
Workflow Roles
Home / Dashboard / Works / Add New Work
Add New Data Set
Descriptions
Files
Relationships
To create a separate work for each of the files, go to Batch upload
Title required
A name to aid in identifying a work.
Add another Title
Creator required
The person or group responsible for the work. Usually this is the author of the content. Personal names should be entered with the last name first, e.g. "Smith, John.".
Add another Creator
Academic Affiliation required
Academic Department/College/School/Unit at CU Boulder
Add another Academic Affiliation
Resource Type required
The general category of this resource (e.g. Masters Thesis, dissertation).
Data Set
Rights Statement required
Additional fields
Save Work
Requirements
Describe your work
Add files
Check deposit agreement
Visibility
Public
Make available to all.
Please note, making something visible to the world (i.e. marking this as Public) may be viewed as publishing which could impact your ability to:
Patent your work
Publish your work in a journal
Check out SHERPA/RoMEO for more information about publisher copyright policies.
Embargo
Set date for future release.
I have read and agree to the Deposit Agreement
Save


Additional fields
Date Issued
The date the resource was published or awarded, such as when an article is published in a journal.
Format: yyyy-mm-dd
Abstract or Summary
A brief description or summary of the item.
File
Edit
View
Format
Formats
POWERED BY TINY
Remove
Subject
Headings or index terms describing what the work is about.
Add another Subject

Example Data Set in Dryad



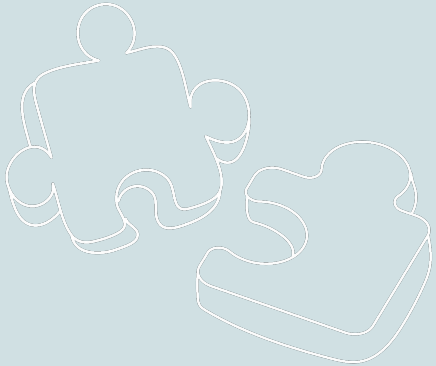
https://doi.org/10.5061/dryad.h18931zpb

CSU & Dryad Resources

- CSU Libraries Data Management website: https://lib.colostate.edu/services/data-management/dryad
- Dryad best practices guide: https://datadryad.org/stash/best_practices
- CSU Open Data guide: https://libguides.colostate.edu/openaccess/opendata

Thank you!

Comments? Questions?



General Questions:
crdds@colorado.edu

CU Scholar Questions:
cuscholaradmin@colorado.edu

CSU & Dryad Questions:
mara.sedlins@colostate.edu