
EPSCoR Research Workspace Demo

And

Ask me anything (data management related)

— Adrienne Canino, Axiom Data Science —
February 2023

This presentation: <https://bit.ly/3YD2hTH>

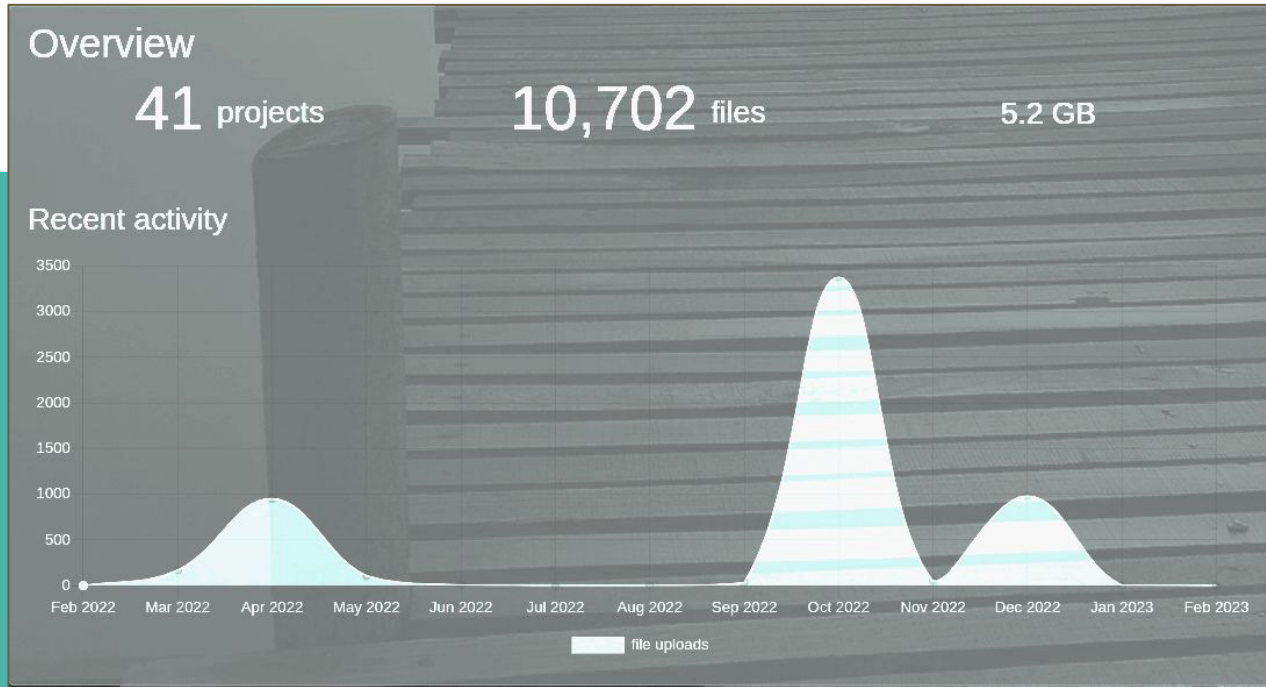
Introducing Axiom and Research Workspace

- Axiom operates a high performance computing center that runs the portal and the Research Workspace
 - The Research Workspace is a cloud platform that provides space for file upload, folders/directory structure, and authoring metadata.
 - Jupyter notebooks are available in Research Workspace for shared data exploration and workflows.
 - The end results will be archived with DataONE or a discipline specific repository, with DOI minted for the dataset.
- Axiom also provides research program support for data planning, management and coordinating (that's me).

This presentation: <https://bit.ly/3YD2hTH>

State of the Data 2023

Summary: Hooray for progress, there's more than 3,000 more files uploaded ! 🎉

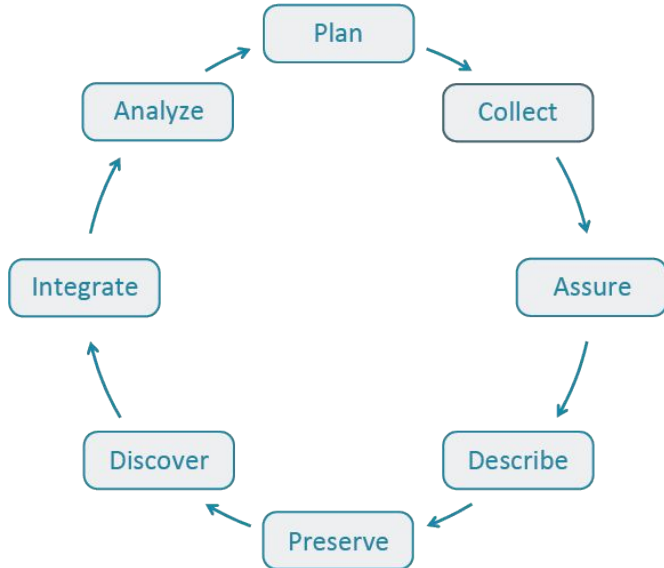


A Quick Review of Data Lifecycles

A loop, connected and interconnected as a web is, that represents the stages data go through from research to repository and reuse, roughly encompassing the concepts of:

plan, collect, quality control, description, preservation, discovery, analysis

DataONE









US FWS



Data Governance plans for EPSCOR Fire and Ice

1. Each PI should have a Research Workspace account and data management plan.
2. Within 6 months of identified field season, preliminary files are in the RW, and preliminary metadata are started
3. Within 12 months, datasets are finalized, metadata are finalized and imported to EPSCoR metadata catalog
4. Final datasets are reviewed by Axiom as part of final QA/QC, before an embargo period of up to 18 months.
5. DOIs are minted for each discrete data set.

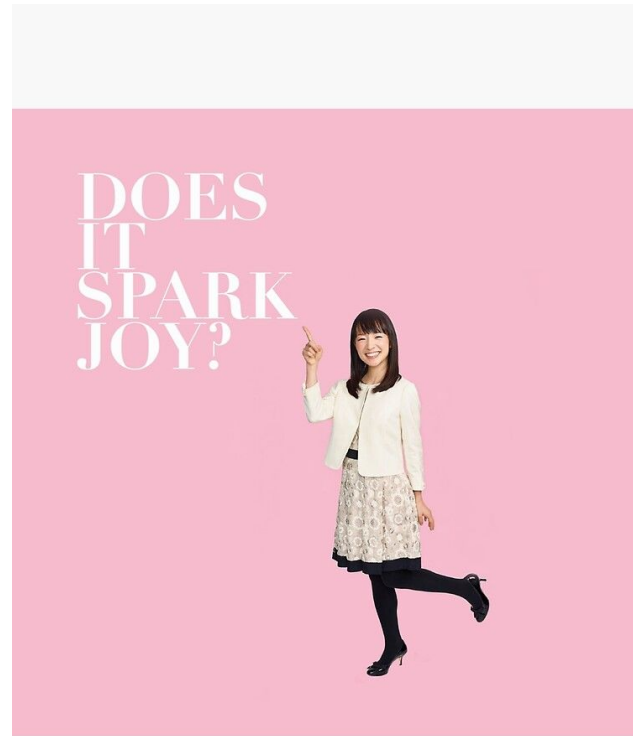
Jargon busting

Portal 	A portal is an organized, searchable tool for finding data and metadata. Typically a portal allows users to perform in-portal data analysis and data visualization such as through a web browser. The portal includes a catalog and a map.
Published 	Any data that is available for download in the Fire & Ice portal is considered published data. These data files are assumed to have been reviewed for completeness and accuracy (QA/QC) by their contributor (e.g. project PI), authorized for publication through a written agreement, and described by an ISO-compliant metadata record. Note that metadata for data files that are currently under embargo may be displayed in the Fire & Ice data portal, though data files themselves will not be available for download until the embargo period ends.
Archived 	Final data and data products are stored in a data repository for long-term preservation. Versioning possible.
Metadata 	Metadata is the data about the data, and describes the data temporally, geographically, data collection and analytical methods, and data contributors in detail. The record in RW is a gray 4 box icon next to a folder name. 
Research Workspace (RW) 	A secure, web-based data management tool used to store and share Fire & Ice data and other associated information for exclusive access by Fire & Ice PIs and their designated collaborators.

An accurate representation of data management

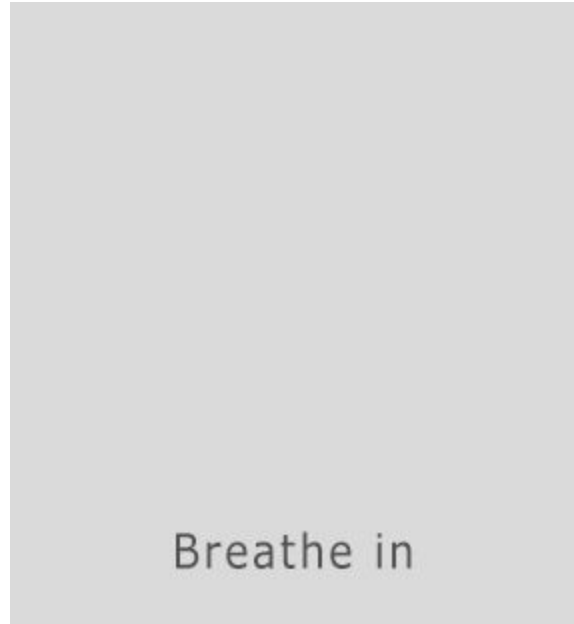


How to do the data management-ing?



This presentation: <https://bit.ly/3YD2hTH>

Time to learn Research Workspace



This presentation: <https://bit.ly/3YD2hTH>

Research Workspace - an active data management platform

Research Workspace is:

- Cloud storage and online collaboration platform
- Project space with multiple permissions levels
- Metadata authoring (ISO 19115 standard, XML export)
- a pipeline to archive with DataONE data repository
- A pipeline to publish to the EPSCOR portal
- Pipeline to mint DOIs

The screenshot displays the Research Workspace interface for the project 'AK NSF EPSCoR Fire & Ice 2018 - 2023'. The interface includes a search bar, navigation tabs (About, Projects, Settings), and a user profile for Adrienne Canino. A left sidebar shows a 'Labels' list with categories like 'administration', 'boreal', 'coastal', 'estuary', 'hyperspectral', 'outreach', and 'seed grant'. The main content area lists several project entries with their dates and file counts, such as 'Administration - Fire & Ice' (Nov 09, 2020, 24 files) and 'BF - Vegetation sites field data and products' (Jun 15, 94 files). A right sidebar titled 'Latest activity' shows recent updates from users like Adrienne and Brenda, including file updates and metadata additions.

Prepping data for sharing in 6 steps

1. Sign into to Research Workspace
2. Build folders for each unique bundle of data and files that should get a single citation.
 - a. Give those folders descriptive names
3. Upload files into those folders.
 - a. Give those files descriptive names.



A note on data formats, with examples

- .xlsx
- .docx
- “All Vemco products”

- .csv
- .odf
- .shp
- .TIFF

- Tidy tables
- ERDDAP tables
- QGS project file

- netCDF
- ACDD
- Darwin Core

- .rproj
- Github or similar repo

**Proprietary
formats**

**Open
formats**

**Open,
standardized
formats**

**+
self-describin
g**

**+ actionable
code, data, +
metadata**

Bad

Acceptable

Achievable

Awesome

Ideal

Prepping data for sharing in 6 steps

4. Write metadata.

a. Preferably at least 2 kinds.

5. Give to Adrienne or Karina for review.

Other team members too.

6. Verify citation, abstract, and number of files in data publishing agreement.



A note on Metadata, with examples

- Field notes
- Proposals
- Papers
- reports

- READMEs
- Data papers

- CSDGM
- ISO
- EML
(on their own)

- Linked keywords
- ORCIDs
- RORs

- WholeTale
- Well-formed GitHub repositories

Ad hoc narrative

Structured narrative

Defined & standard format

+ defined semantics

+ actionable code, data, + metadata

Bad

Weak, useful

Acceptable

Achievable

Ideal

Done



A **DOI** is minted.

A data set is open and free.

Example dataset: Northern Seadragon Migrations

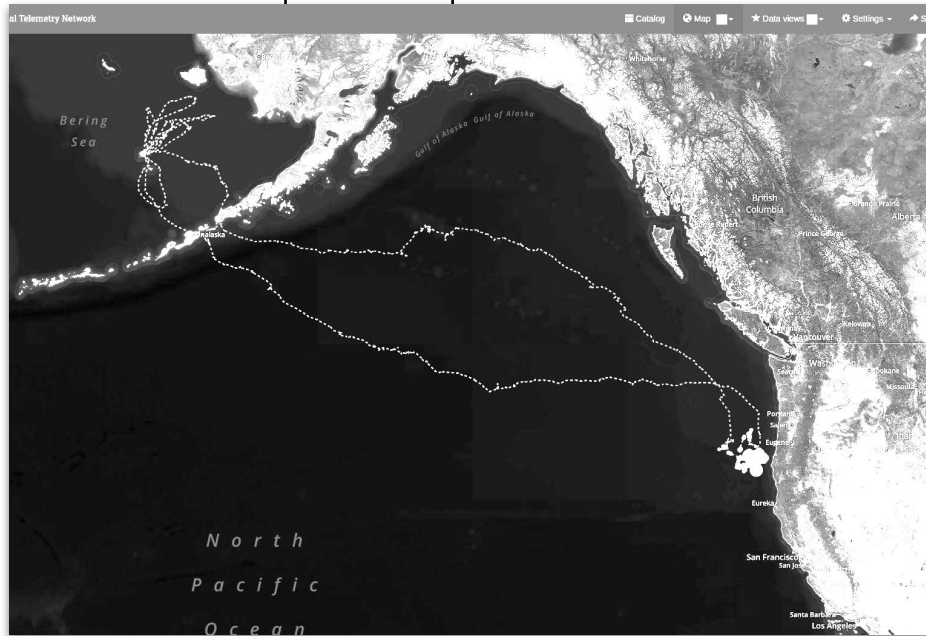
RW link: <https://bit.ly/3nXZXWs>

This dataset is presented here as 1
CSV file:
northern-seadragons-migration-data-
2009-2010.csv.



Example data

ptt-id	location-datetime-utc	location-latitude	location-longitude	location-quality
10101	10/03/09 03:24 AM	57.162	-170.159	1
10101	10/03/09 04:08 AM	57.157	-170.255	0



57.157	-170.199	0
57.177	-170.156	A
57.111	-170.216	A
57.158	-170.176	1
57.162	-170.177	3
57.162	-170.175	3
57.161	-170.173	3

Example data: Northern Seadragon Migrations

Project Abstract:

This dataset is composed of location and location quality data from satellite tags attached to two north pacific seadragons. These data are from the Bering Sea and North Pacific Ocean, ranging from St. Paul Island to the Oregon Coast. These data were collected by satellite tags attached to the seadragons that transmitted location data when available through the Argos satellite network. After the end of the deployments, data was downloaded through the Wildlife Computers data portal and processed as described in the Process Steps field of the Methods section of this metadata record.

On the portal: <https://ak-epscor.portal.axds.co/#>

Example Dataset Live Demo

- Uploading files
- Rearranging files
- Re-naming files
- New Folders
- New metadata
- NOT FILE LEVEL METADATA
- Copying a template for metadata
- The portal view
- The DOI view (DataONE)
- The Jupyter notebook

Jupyter Notebooks Demo

https://researchworkspace.com/file/39372791/seadragon_telemtry_notebook.ipynb

What are Jupyter notebooks?

- IDE-like coding environment
- Mix code and documentation
 - Hooray for ABD!
- Browser-based
 - Hooray for anywhere access!
- Built in to the Research Workspace, so it can see your files right next to the notebook

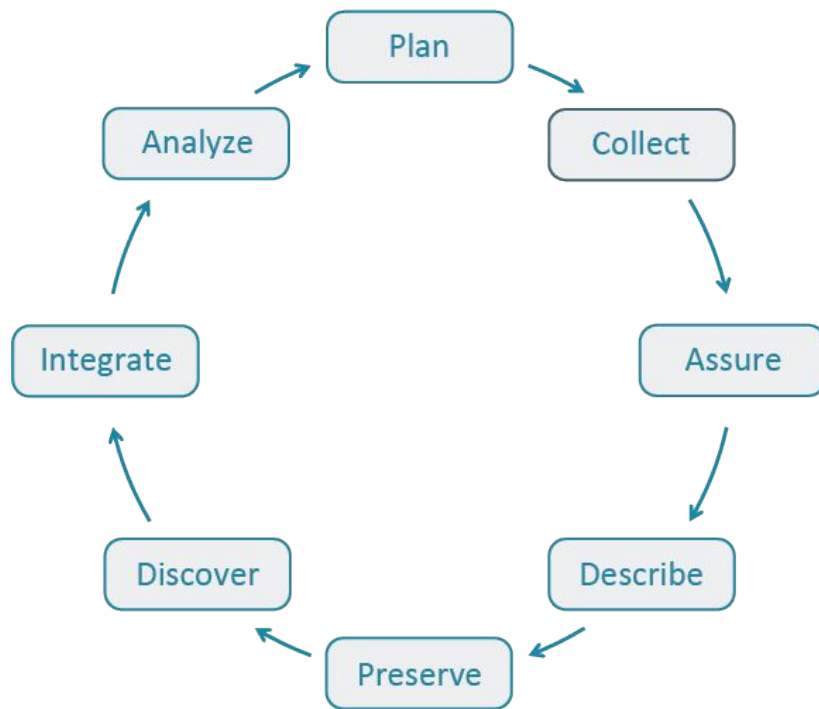
Link Soup, aka Helpful resources

- Using the Research Workspace:
<https://researchworkspace.com/help/NavigatingTheWorkspaceTutorial.html>
- Metadata Guide:
<https://www.axiomdatascience.com/best-practices/MetadataBestPractices.html>
- Metadata Cheat Sheet:
<https://www.axiomdatascience.com/best-practices/MetadataCheatSheet.html>
- This example project in the Research Workspace:
<https://researchworkspace.com/project/39372508/folder/39372510/data> (may need permissions)
- My favorite README:
https://social-science-data-editors.github.io/template_README/template_README.html
- The Epscor data portal:
<https://ak-epscor.portal.axds.co/#>

This presentation: <https://bit.ly/3YD2hTH>

Data Management Learning Opportunities

- [DataONE Best Practices](#)
- [DataONE Data Management Skillbuilding Hub](#)
- [Data Management Training Clearinghouse](#)
- [Data Carpentry](#) - Domain specific data mgmt training
- [Software Carpentry](#) - Basic skills for research computing





Now we're ready.

Adrienne:
adrienne@axiomdatascience.com

Thank You!

Questions?