Introduction to Data Management and Publication

By Gabe Kamener
FCE Information Manager, Florida International University

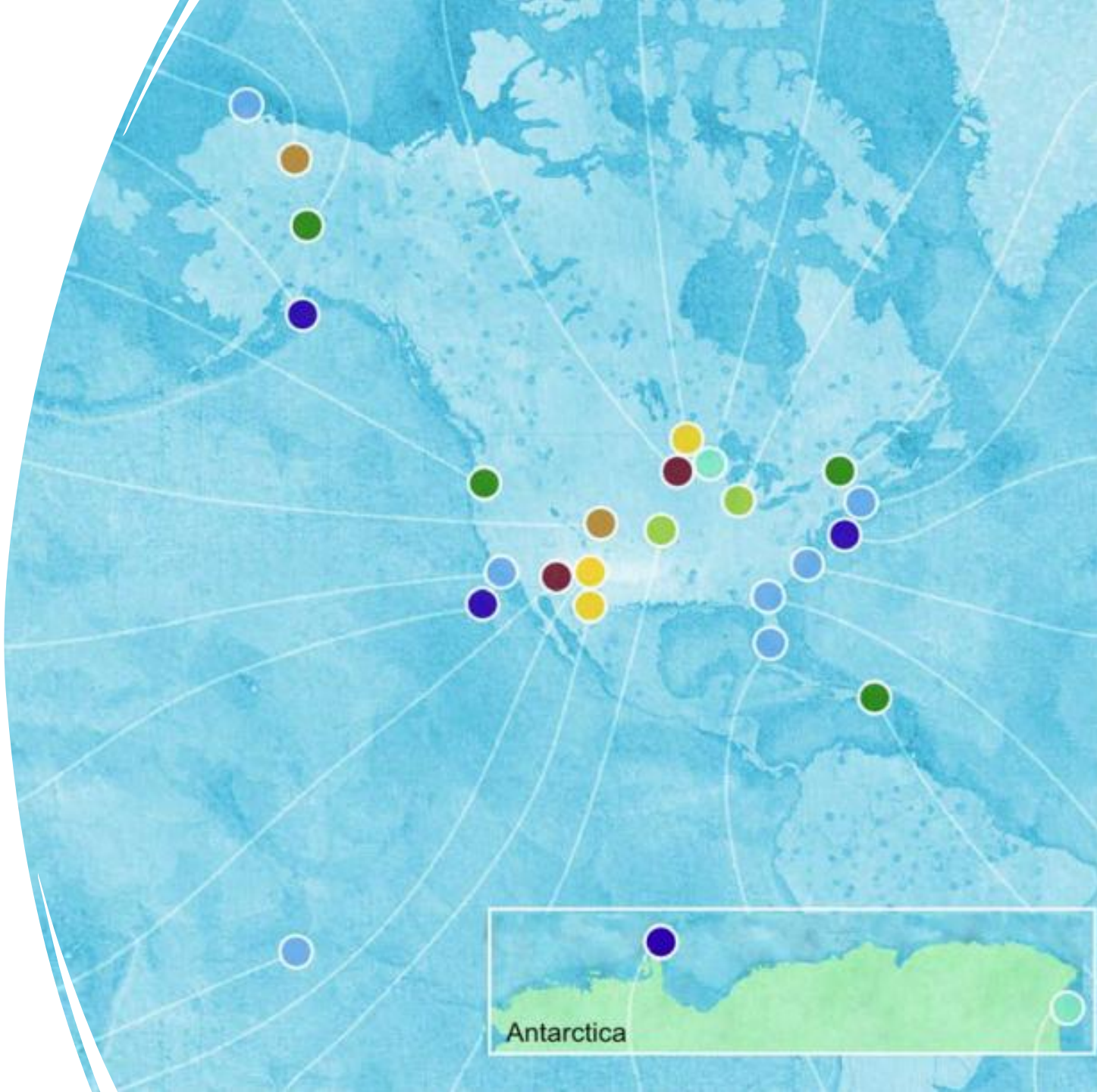
2026 Florida Coastal Everglades LTER Information Management
January 22, 2026

Overview

- Review
  - Sharing of LTER Network data
  - EDI data repository and FCE data catalog
  - Citing data
- Best practices
  - Data and project management
  - Formatting data
  - Describing metadata
- Preparing for publication with ezEML

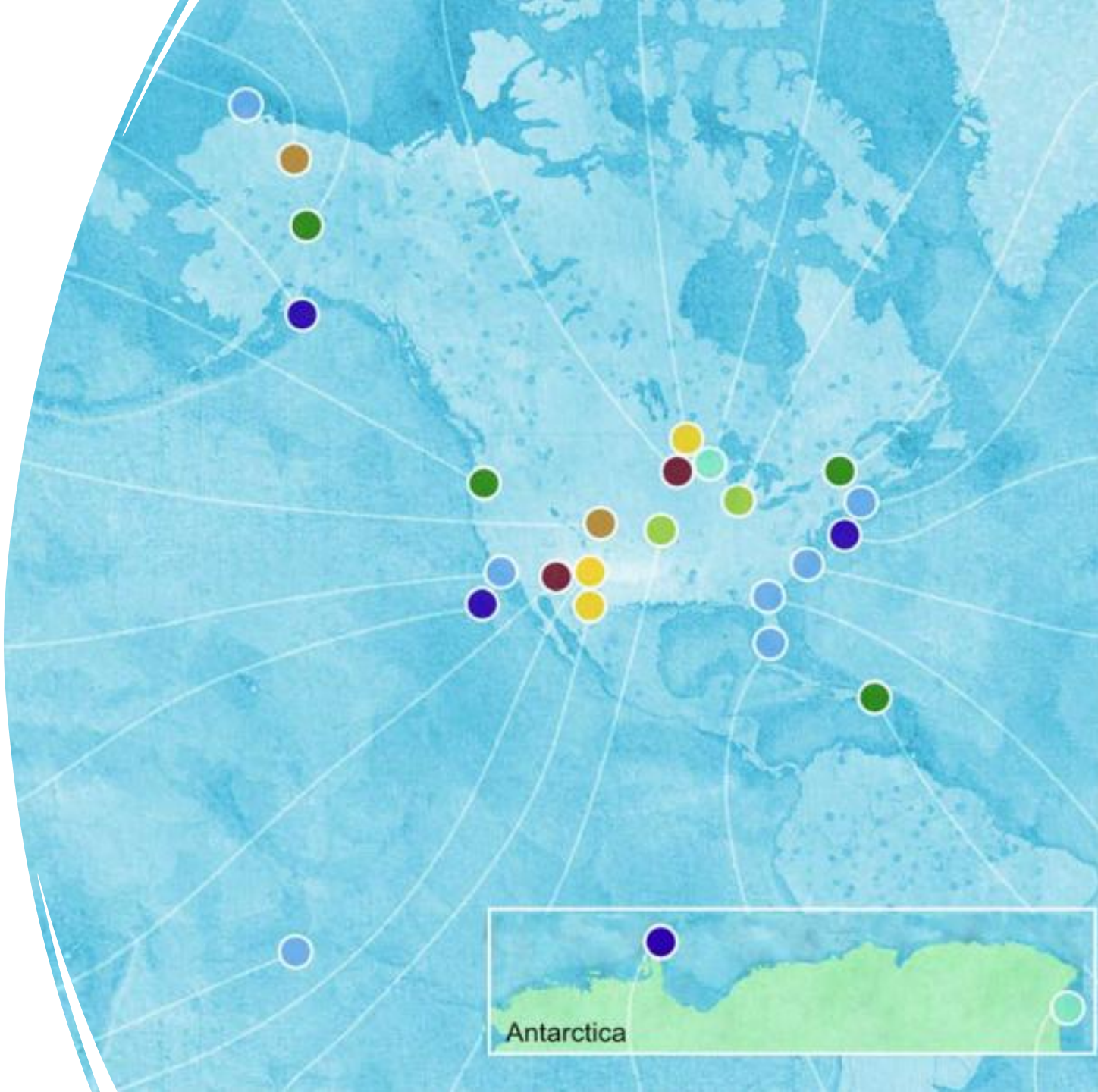
Sharing of LTER Network Data

- Ongoing since NSF funded first LTER sites in 1980
- Enables new science!
- Supports open science and reproducibility
- Funders and journal publishers often require datasets in a recognized data repository



LTER Network Data Release Policy

- Data and information derived from publicly funded research in the U.S. LTER Network, totally or partially from LTER funds from NSF… [must be].. made available in a community accepted data repository… with as few restrictions as possible, on a nondiscriminatory basis.



Two Types of LTER Data

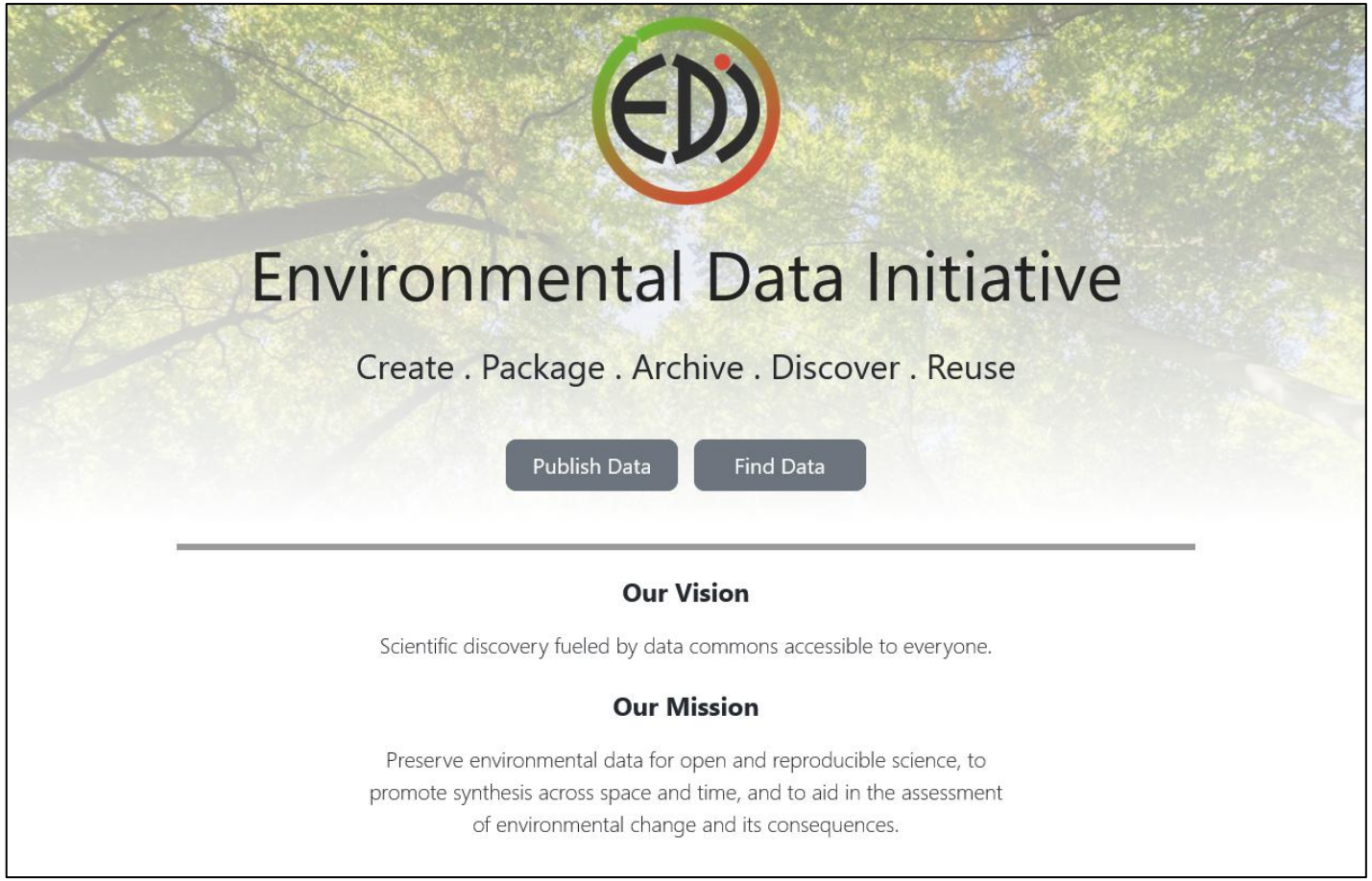
- Type I – data are to be released to the general public according to the terms of the general data use agreement within 2 years from collection and no later than the publication of the main findings from the dataset
- Type II - data are to be released to restricted audiences according to terms specified by the owners of the data. Type II data are considered to be exceptional and should be rare in occurrence.

FCE LTER Graduate Student Data

- Submit complete dataset and metadata to FCE Information Manager (IM) before graduation
- Meet with IM **at least 2 months before** you need to publish data

The Environmental Data Initiative (EDI) Data Repository

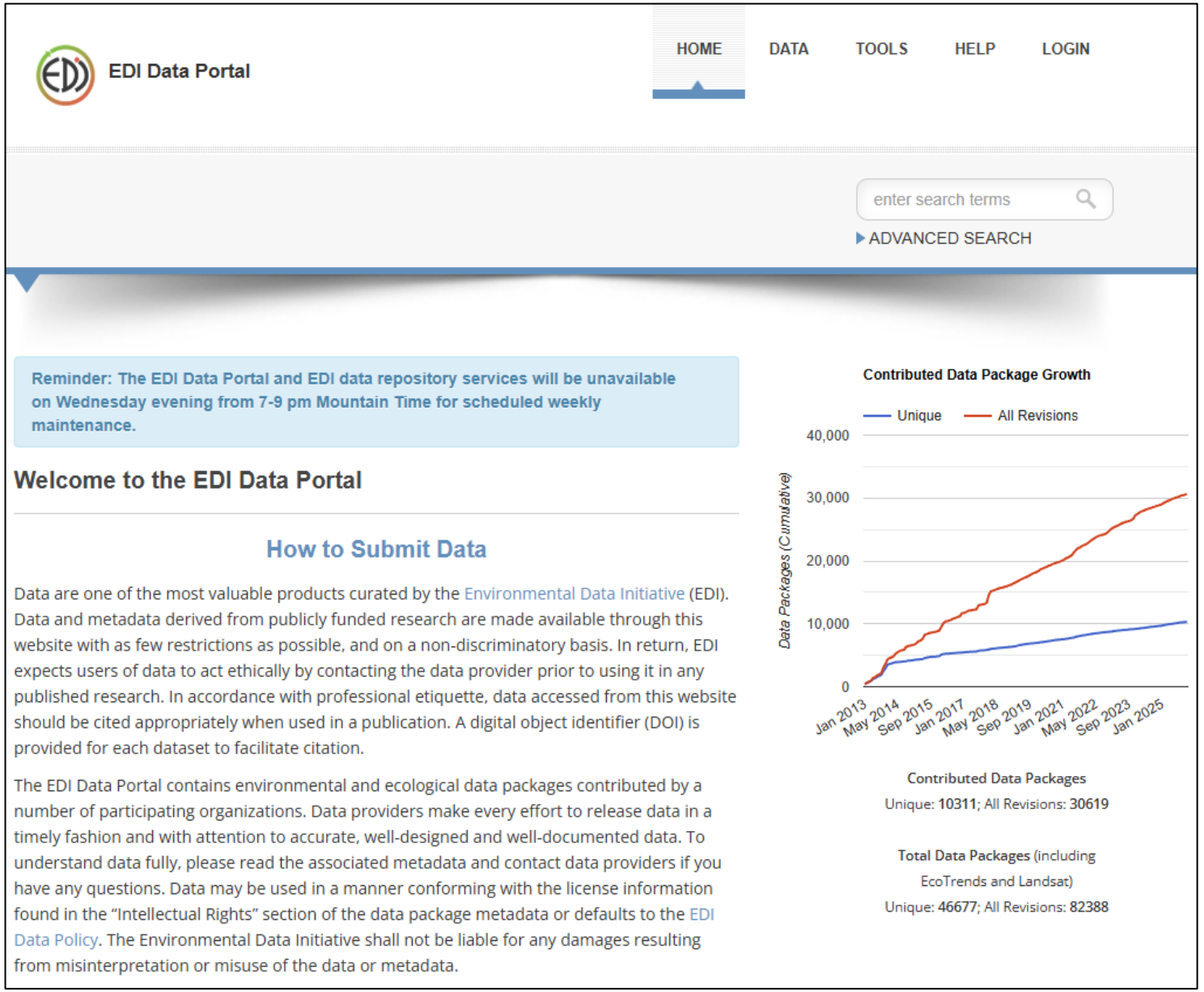
- Funded by NSF
- Provides:
  - Long-term data security
  - Long-term data accessibility
  - Data integrity
  - Data discovery
  - Citable digital object identifiers (DOIs) for datasets



https://edirepository.org

EDI Data Portal

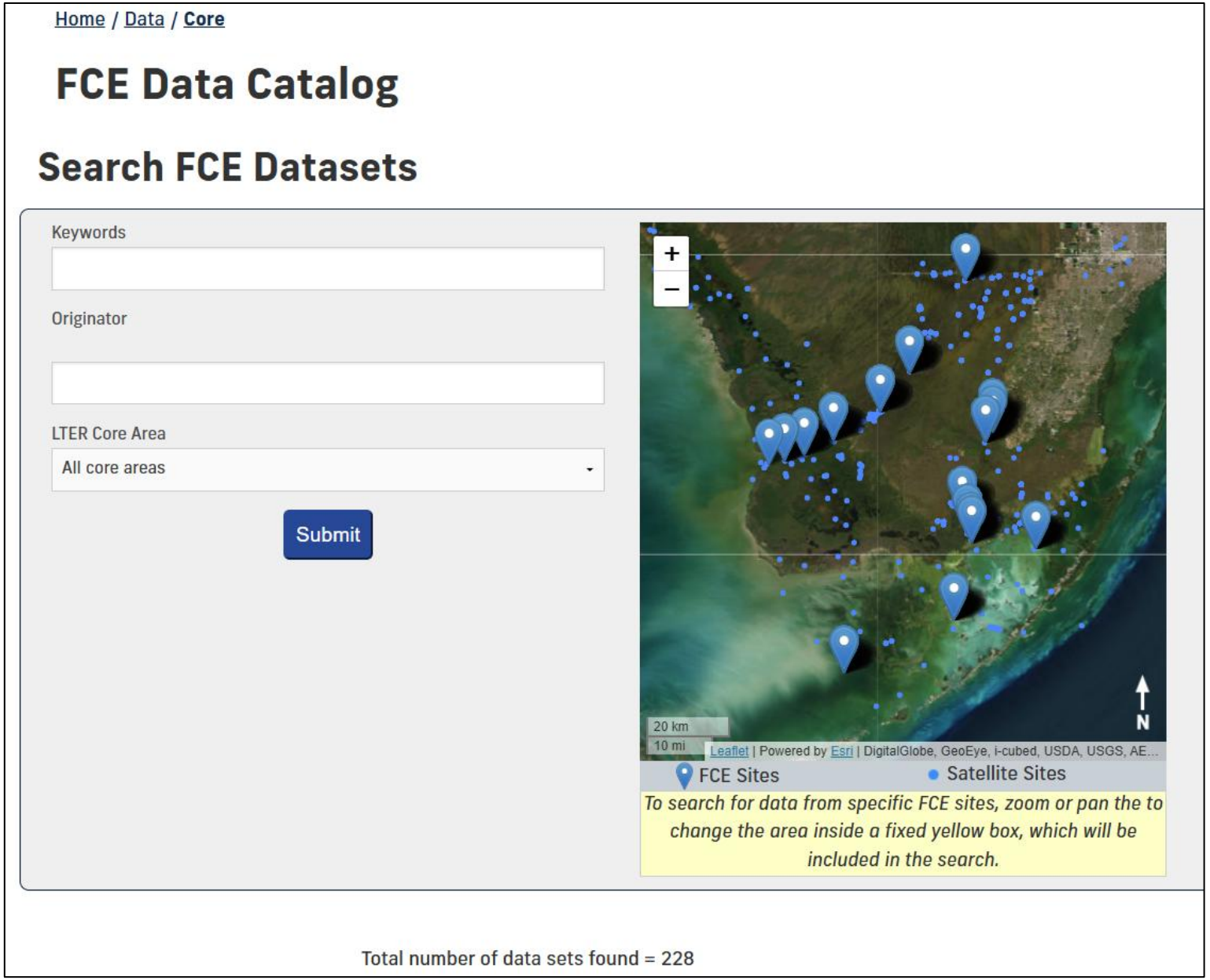
- User-friendly interface of EDI repository
- More than 10,300 searchable, unique data packages
- Advanced search functionality



https://portal.edirepository.org/nis/home.jsp

FCE Data Catalog

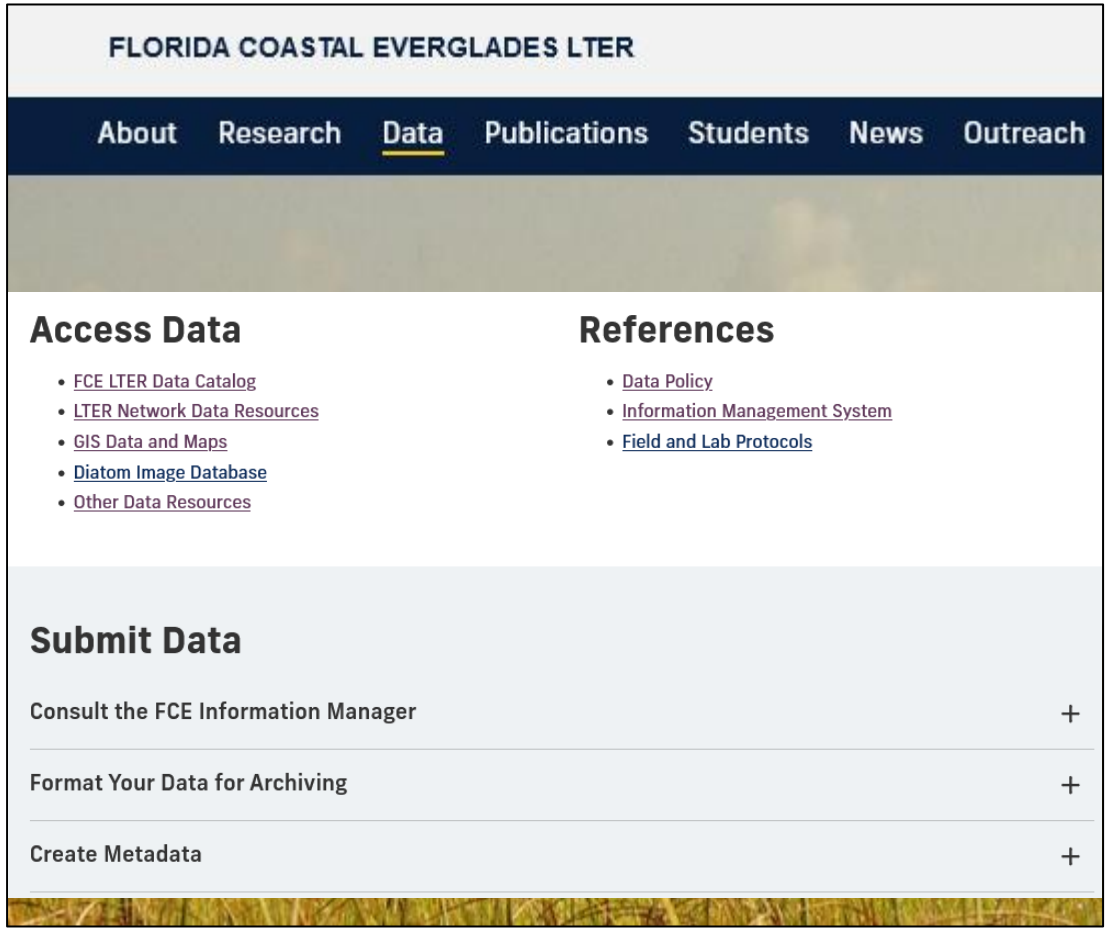
- Lists FCE datasets published in EDI
- Easily search the catalog from the FCE website!



https://fce-lter.fiu.edu/data/core/index.php

Steps to Publish Data in the EDI Repository

1. Review the FCE Data page!
2. Contact the FCE IM about:
   - Data (tabular, model code, imagery, etc.)
   - Required metadata
   - How to format your dataset
   - Getting an FCE dataset ID
3. Enter data and metadata into ezEML
4. Review package with FCE IM and publish!



https://fcelter.fiu.edu/data

Citing Data

Kleindl, P.M., E. Gaiser, A. Wachnicka, J. Sah, and M. Ross. 2024. Macrophyte and microbial mat biomass co-variation along a hydrologic gradient and response to a removal experiment in temporary wetlands Everglades, FL, USA, February 2003 – November 2006 ver 3. Environmental Data Initiative. https://doi.org/10.6073/pasta/01ef607bc7090c3488e5bcb7475847e9 (Accessed 2026-01-22)

Title: Macrophyte and microbial mat biomass co-variation along a hydrologic gradient and response to a removal experiment in temporary wetlands Everglades, FL, USA, February 2003 – November 2006

Creators: Kleindl, Paige Marie; Ph.D. Candidate - Periphyton Laboratory; Florida Coastal Everglades LTER
Gaiser, Evelyn; Principal Investigator; Florida Coastal Everglades LTER
Wachnicka, Anna; Principal Scientist; South Florida Water Management District…

Show more >

Publication Date: 2024-12-04

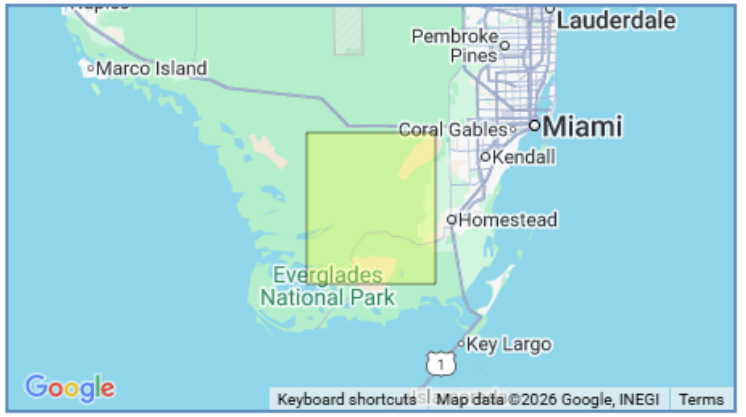
Citation: Kleindl, P.M., E. Gaiser, A. Wachnicka, J. Sah, and M. Ross. 2024. Macrophyte and microbial mat biomass co-variation along a hydrologic gradient and response to a removal experiment in temporary wetlands Everglades, FL, USA, February 2003 – November 2006 ver 3. Environmental Data Initiative. https://doi.org/10.6073/pasta/01ef607bc7090c3488e5bcb7475847e9 (Accessed 2026-01-22).

Copy Citation

Abstract: This data package encompasses hydrologic variables, soil depth, hydrologically-regulated macrophyte community types, macrophyte biomass and community structure, and microbial mat biomass that was collected in two observational surveys and one in-situ experimental manipulation in six temporary wetland regions located in the…

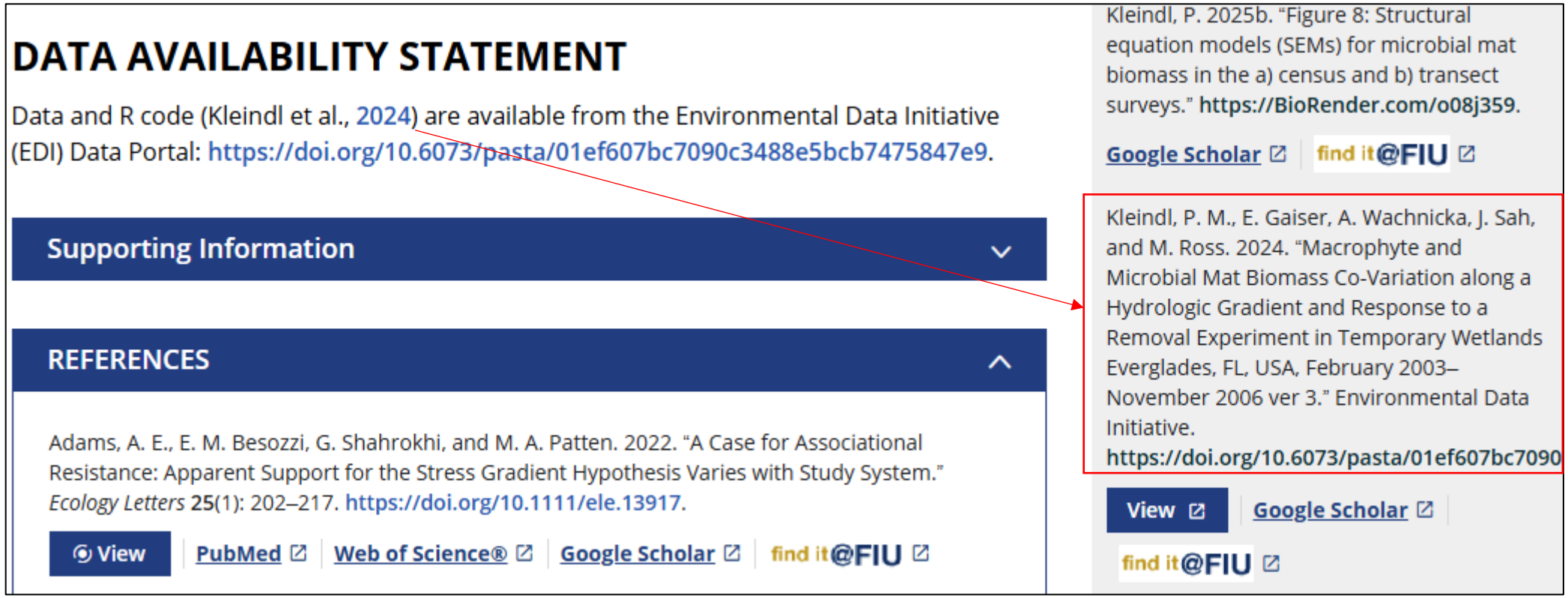
Show more >

Spatial Coverage:



Citing Data

- Include full citations (with DOI) in References and then cite those references in Data Availability Statement (also include DOI links there)



Citation use in Kleindl et al. (2025). https://doi.org/10.1002/ecs2.70384

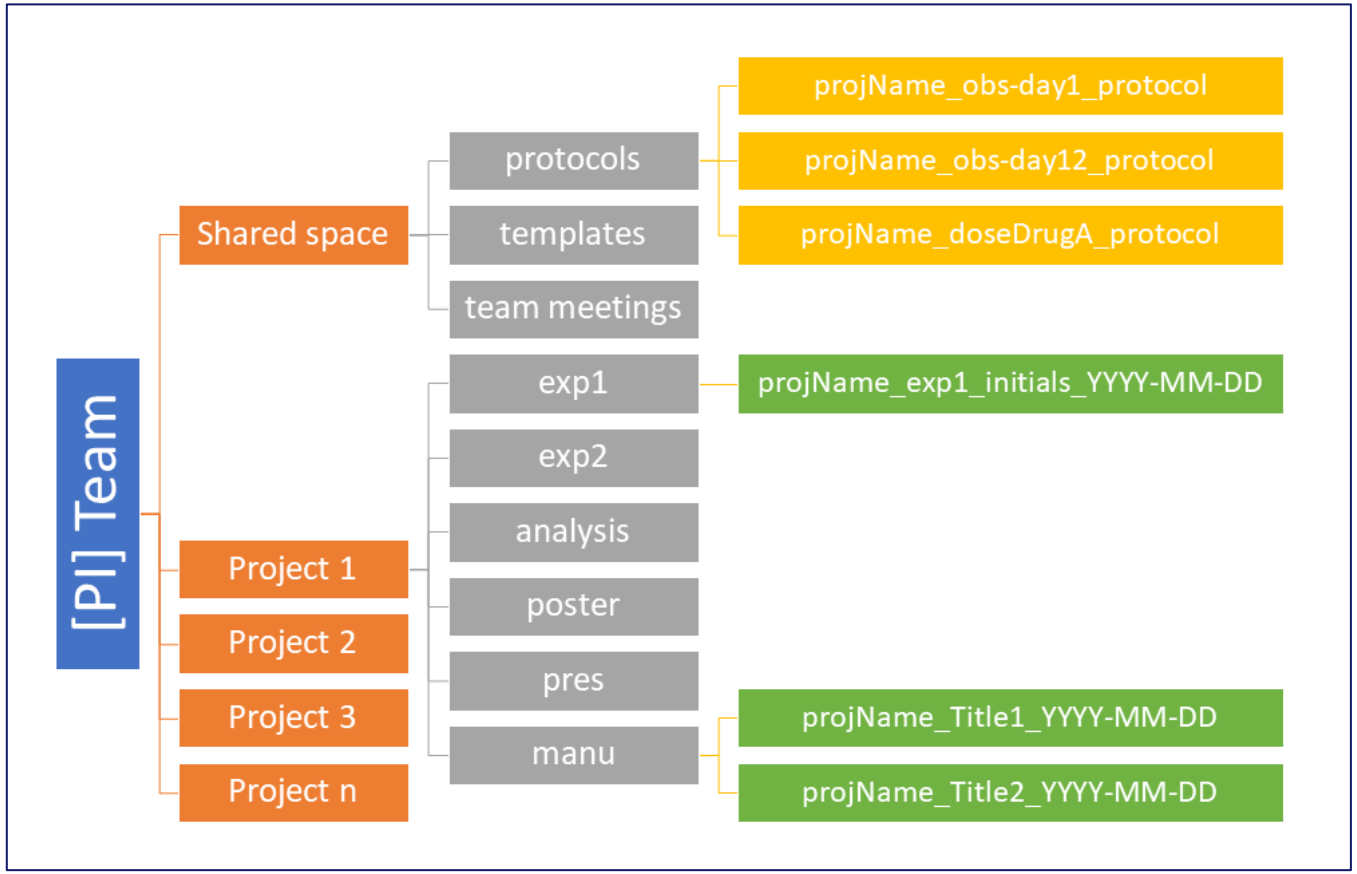
Best Practices in Data Management

“Well managed data is a benefit to any researcher as it requires less digging to find, less effort to understand, and less processing to prepare for collaboration, reuse, and sharing.” -Briney et al. 2020.

Artwork by @allison_horst (CC BY 4.0)

File Organization

- Use folder structures
- Use consistent file names
- Keep raw data separate from analysis
- Use file versioning



Briney et al. 2020

Backup Your Data

- 3-2-1 rule
  - **Three** copies of the data
  - **Two** geographically separate locations
  - More than **one** type of storage device

Write a Living Data Management Plan

- Document important details (e.g. file organization and storage, backup plan, etc.) in one place
- Can be relatively short
- Update as research project evolves

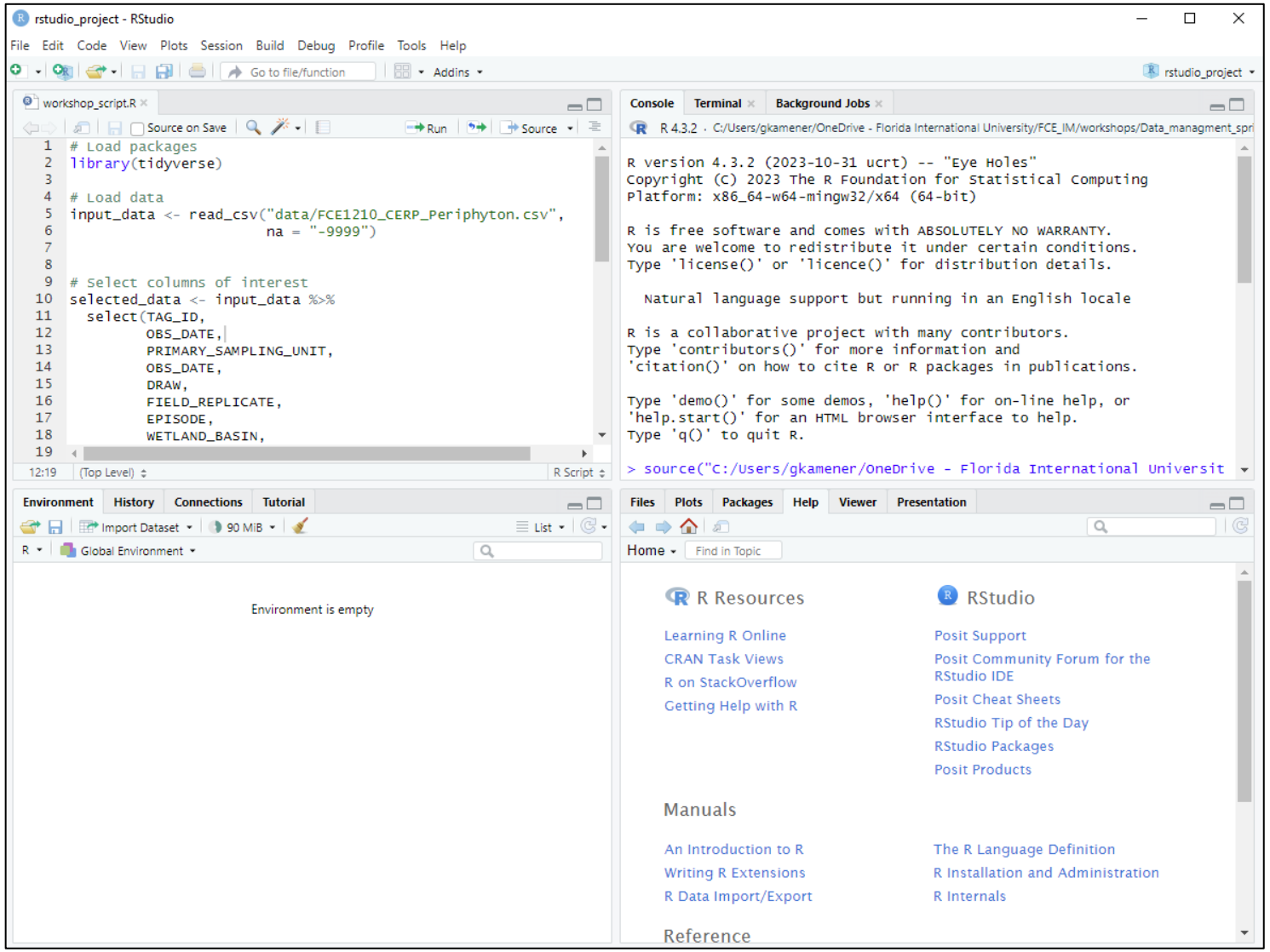
Artwork by @allison_horst (CC BY 4.0)

Sharpen Your Data Management Skills

- Briney, K. A., Coates, H. L., & Goben, A. (2020). Foundational practices of research data management. Research Ideas and Outcomes 6: e56508. https://doi.org/10.3897/rio.6.e56508
- Briney, K. (2023). The Research Data Management Workbook. Caltech Library. https://doi.org/10.7907/z6czh-7zx60

Briney 2023

Project Management in RStudio



RStudio Projects

Advantages of RStudio Projects:

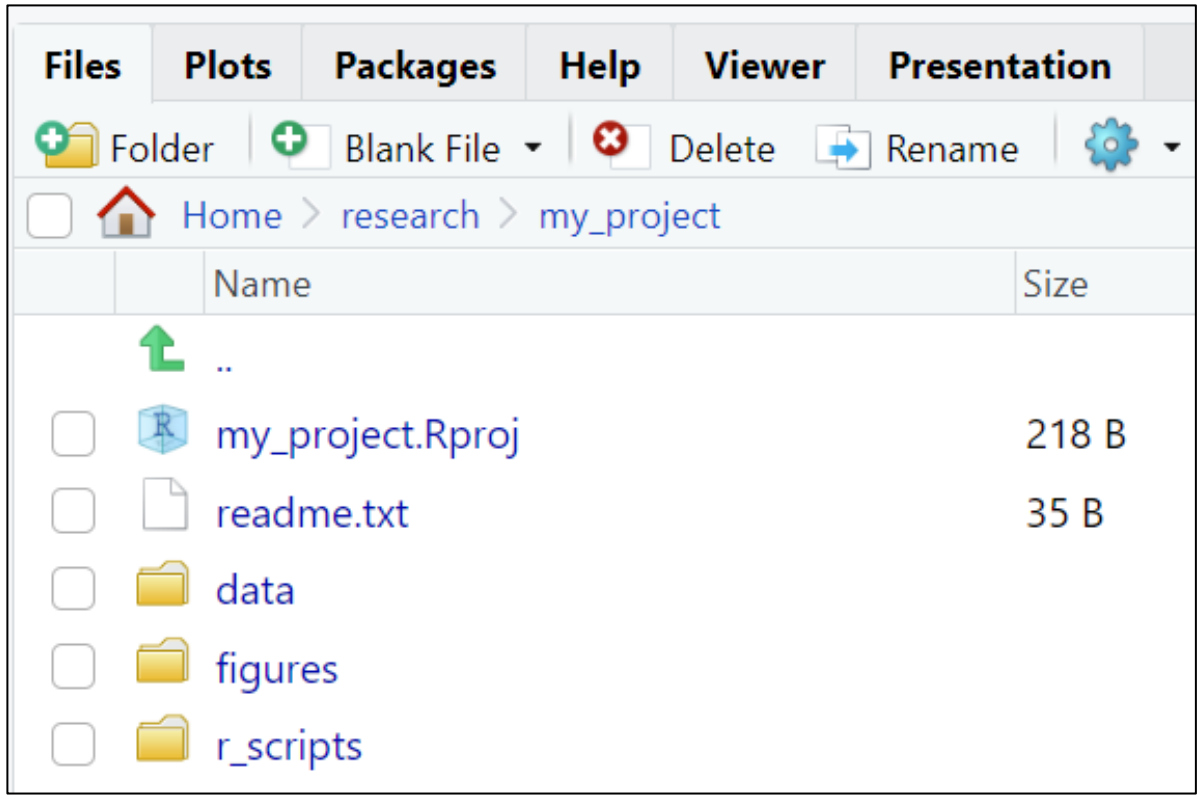
- **Automatically set working directory**
- **Portable**
- **Collaborator friendly**
- Supports version control with Git/GitHub
- Can aid reproducibility with renv package
- **Can integrate with GitHub Copilot**

Setting up an RStudio Project

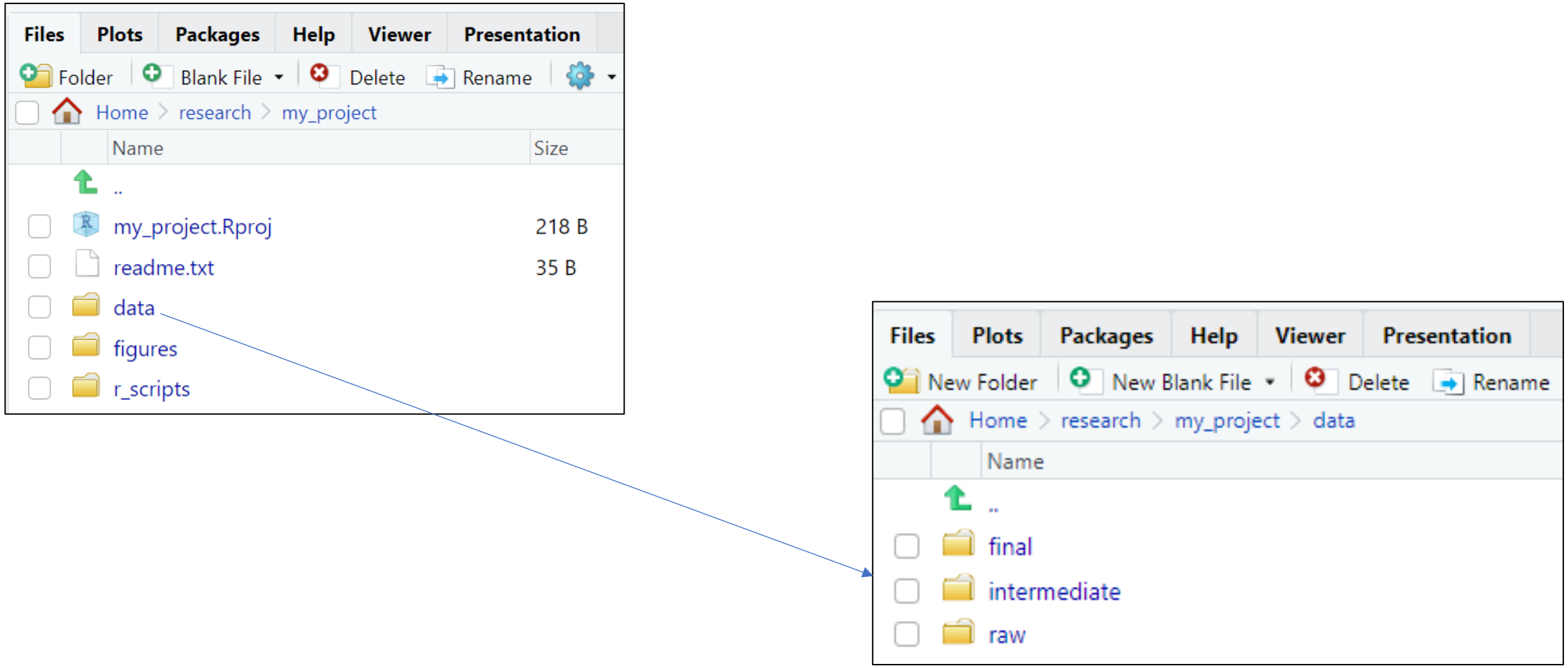
1. Start RStudio.
2. Under the File menu, click on New Project. Choose **New Directory**, then **New Project**.
3. Enter a name for this new folder (or “directory”) and choose a convenient location for it. This will be your working directory.
4. Click on Create Project.
5. Create folders for data (with “raw”, “intermediate”, and “final” subfolders), r scripts, and figures in your working directory.
6. Place raw data files and script files into respective folders.

Adapted from: https://datacarpentry.github.io/R-ecology-lesson/introduction-r-rstudio.html#getting-set-up-in-rstudio

Organizing Your Working Directory



Organizing Your Working Directory



RStudio Projects: Portability

- R script without RStudio project:

```
df <- read_csv("C:/Users/gkamener/Documents/research/my_project/
                data/raw/my_project_data_GK_2024_11_13.csv")
```

  - Often requires specifying working directory or full file path
  - Script may break if “my_project” folder is moved or shared

- With RStudio project:

```
df <- read_csv("data/raw/my_project_data_GK_2024_11_13.csv")
```

  - Path always relative to “my_project” folder
  - Project becomes portable

GitHub Copilot in RStudio

- “Any use of generative ‘AI’ should be **transparent**, **accountable** and **acknowledged**. Check the editorial policies of journals before submission to ensure you are using the most up-to-date guidance” – Cooper et al. 2025
- Check with your advisor and institution before using generative AI for anything research related.

GitHub Copilot in RStudio

- “You need a fundamental understanding of coding to be able to use Copilot effectively.” – O’Hara and Kui 2025
- Following data-management best practices will enable better use of GitHub Copilot

GitHub Copilot in RStudio

- GenAI programming resources recently featured during LTER IM monthly meeting
  - Casey O'Hara and Li Kui (2025), AI for Ecologists. URL https://nceas-learning-hub.github.io/ai_for_nceas.
  - Elmendorf (2026), NWT LTER AI Tools Introduction. URL https://nwtlter.github.io/data_skills/AI_intro_to_AI.html

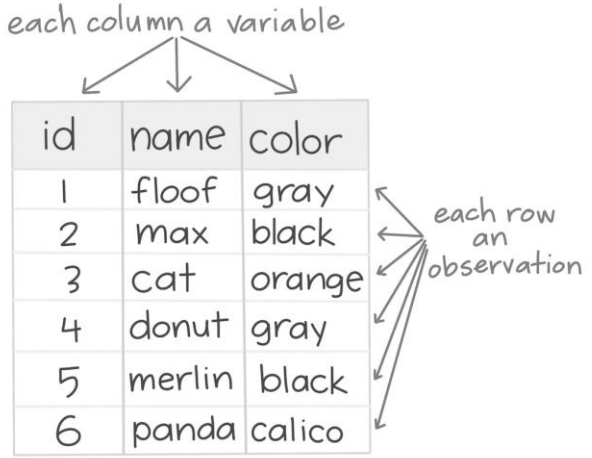
Formatting Data: Tidy is the Goal

"TIDY DATA is a standard way of mapping the meaning of a dataset to its structure."

–HADLEY WICKHAM

In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement



Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

Formatting Data

- One rectangle of data
  - No blank columns or rows
  - No blank cells
- One data type per column
- One value per cell
- Use simple headers
  - No spaces
  - No special characters

Plot #	Subject ID	Weight1207	Stage1207	Sex
1	1	33.2, 1		m
five	2? Not 100% sure	44.0	2	F
#1	3		three	male

Date	Plot	Subject_ID	Weight	Stage	Sex	Comment
2012-07-01	1	1	33.2	1	m	NA
2012-07-01	5	2	44.0	2	f	Unsure of id
2012-07-01	1	3	-9999.0	3	m	Forgot to measure

Formatting Data

- Format dates as YYYY-MM-DD
- Use consistent categorical variable codes
- Avoid calculations or graphs

Plot #	Subject ID	Weight1207	Stage1207	Sex
1	1	33.2, 1		m
five	2? Not 100% sure	44.0	2	F
#1	3		three	male

Date	Plot	Subject_ID	Weight	Stage	Sex	Comment
2012-07-01	1	1	33.2	1	m	NA
2012-07-01	5	2	44.0	2	f	Unsure of id
2012-07-01	1	3	-9999.0	3	m	Forgot to measure

Formatting Data

- Don’t highlight cells or embed comments
- Document any changes you make
- Store metadata in separate sheet or file

Plot #	Subject ID	Weight1207	Stage1207	Sex
1	1	33.2, 1		m
five	2? Not 100% sure	44.0	2	F
#1	3		three	male

Date	Plot	Subject_ID	Weight	Stage	Sex	Comment
2012-07-01	1	1	33.2	1	m	NA
2012-07-01	5	2	44.0	2	f	Unsure of id
2012-07-01	1	3	-9999.0	3	m	Forgot to measure

Metadata

- What are metadata?
  - Metadata = data about data
  - Document **who**, **what**, **why**, **where**, and **when**
- Why use metadata?
  - Keep track of important details about data
  - Required to publish datasets
  - Increase findability and usability of published data

Documenting Metadata

- Dataset metadata should include:
  - Data table information (i.e. data dictionary)
  - Title
  - Abstract
  - Parties responsible for dataset
  - Methods
  - Intellectual rights
  - Keywords
  - Geographic, temporal, and taxonomic coverage
  - Information about non-tabular data (if applicable)
  - Project funding
  - Permits (if applicable)

Documenting Metadata: Data Dictionary

- Describes data table variables (Broman and Woo 2018)
- Create as early as possible!

Column Name	Definition	Variable Type	Units	Precision	Codes	Date Time Format String	Missing Value Code	Missing Value Code Explanation
SITENAME	Name of site	Categorical			SRS2 = Shark River Slough 2			
Date	Date of sample collection	DateTime				YYYY-MM-DD		
Time	Time of sample collection	DateTime				hh:mm	NA	Not recorded
Salinity	Concentration of salinity	Numerical	PSU	0.1			-9999.0	Not recorded
TN	Concentration of total nitrogen	Numerical	micromolePerLiter	0.001			-9999.000	Not recorded
TP	Concentration of total phosphorus	Numerical	micromolePerLiter	0.01			-9999.00	Not recorded
Comment	Field comments	Text					NA	Not recorded

Documenting Metadata

- Title
  - Should be descriptive, including what, where, and when
  - **Do not** use manuscript title
- Abstract
  - Describe who, what, why, where, and when of dataset in more detail
  - **Do not** include results or conclusions from study
- Methods
  - Detail to answer any questions someone might have
  - Include list of cited references (if applicable)
- Intellectual rights
  - FCE uses CC-BY Creative Commons license

Documenting Metadata

- Keywords
  - Source from LTER Controlled Vocabulary* when possible
- Geographic coverage
  - Coordinates in decimal degrees for publishing
- Temporal coverage
  - Documents when data were collected
  - Format as YYYY-MM-DD or YYYY
- Taxonomic coverage
  - Important to check spelling of names

*LTER Controlled Vocabulary link: https://vocab.lternet.edu/vocab/vocab/index.php

Documenting Metadata

- Information about non-tabular data entities
  - Includes metadata about model code, geospatial, imagery, or other data entities
- Project funding
  - Information about funding for data collection
- Permits
  - Permits that were required for data collection

Timing and Importance of Quality Metadata

- Important for future you!
- Important for discovery and re-use of published data!
- LTER Network and EDI use Ecological Metadata Language (EML) to document quality metadata

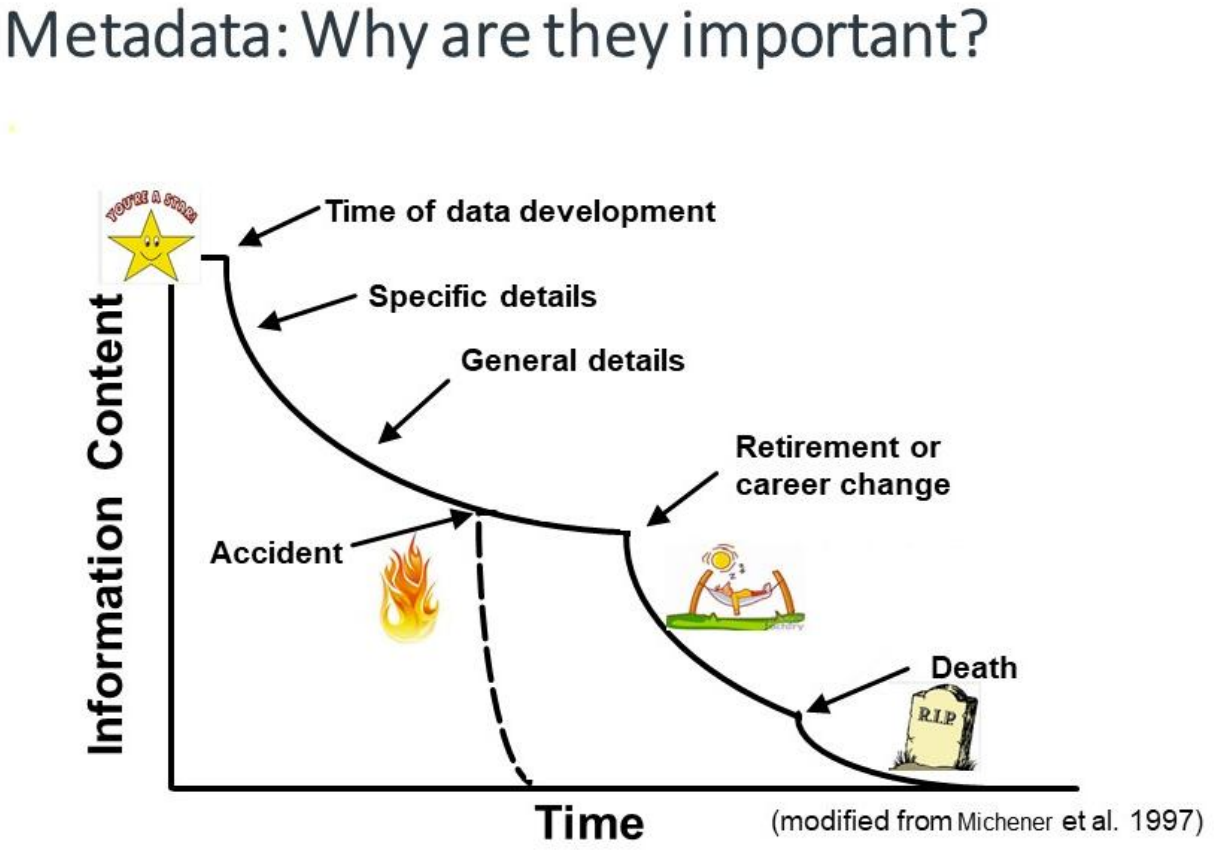


Image courtesy of DataONE

Metadata Creation During Research Life Cycle

- Metadata should be described throughout the research life cycle
  - Describe as much as possible in early stages
  - Regularly add/update throughout life cycle

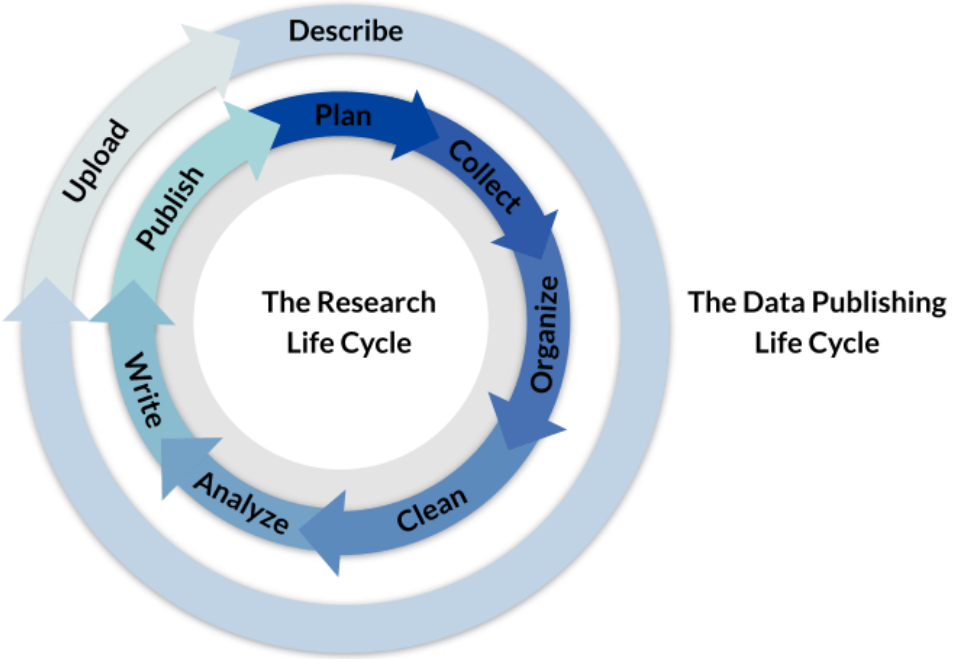
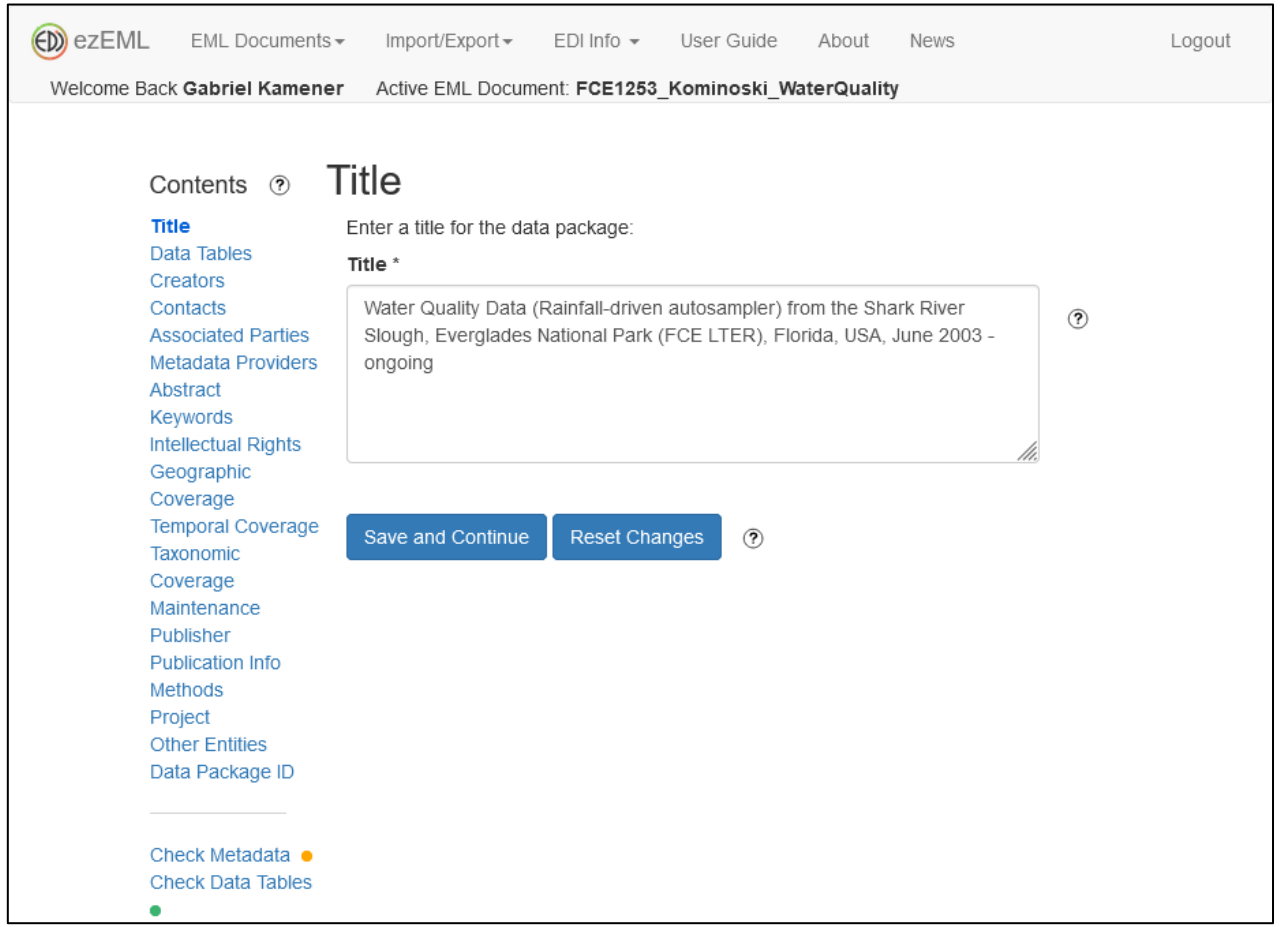


Image courtesy of Environmental data Initiative

ezEML's Role in Data Publishing Life Cycle

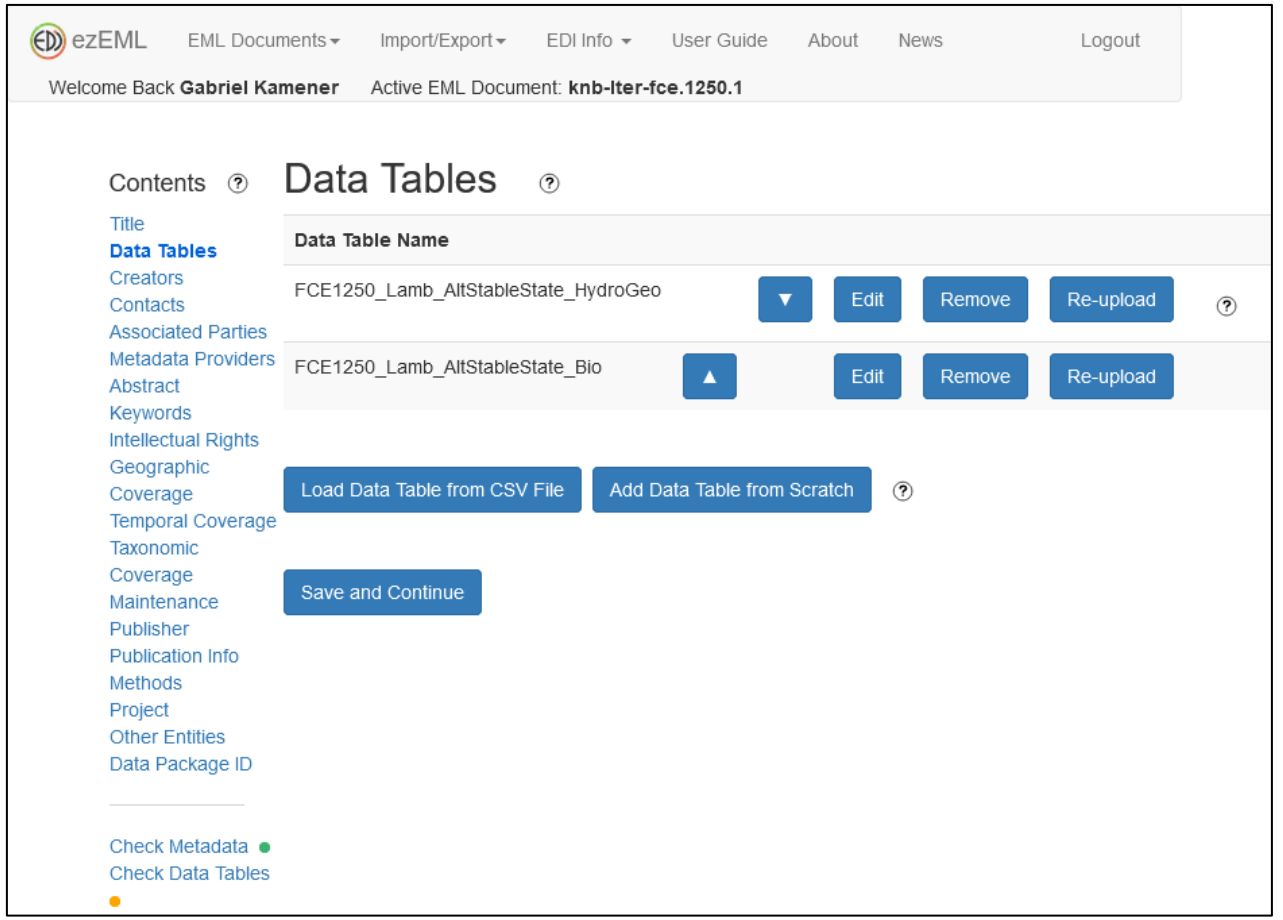
- Form-based web interface to create and edit structured metadata in EML
- Can utilize pre-filled FCE templates
- Saves as you go
- Can quality check data and metadata
- Produces data package (data + metadata)



https://ezeml.edirepository.org/eml

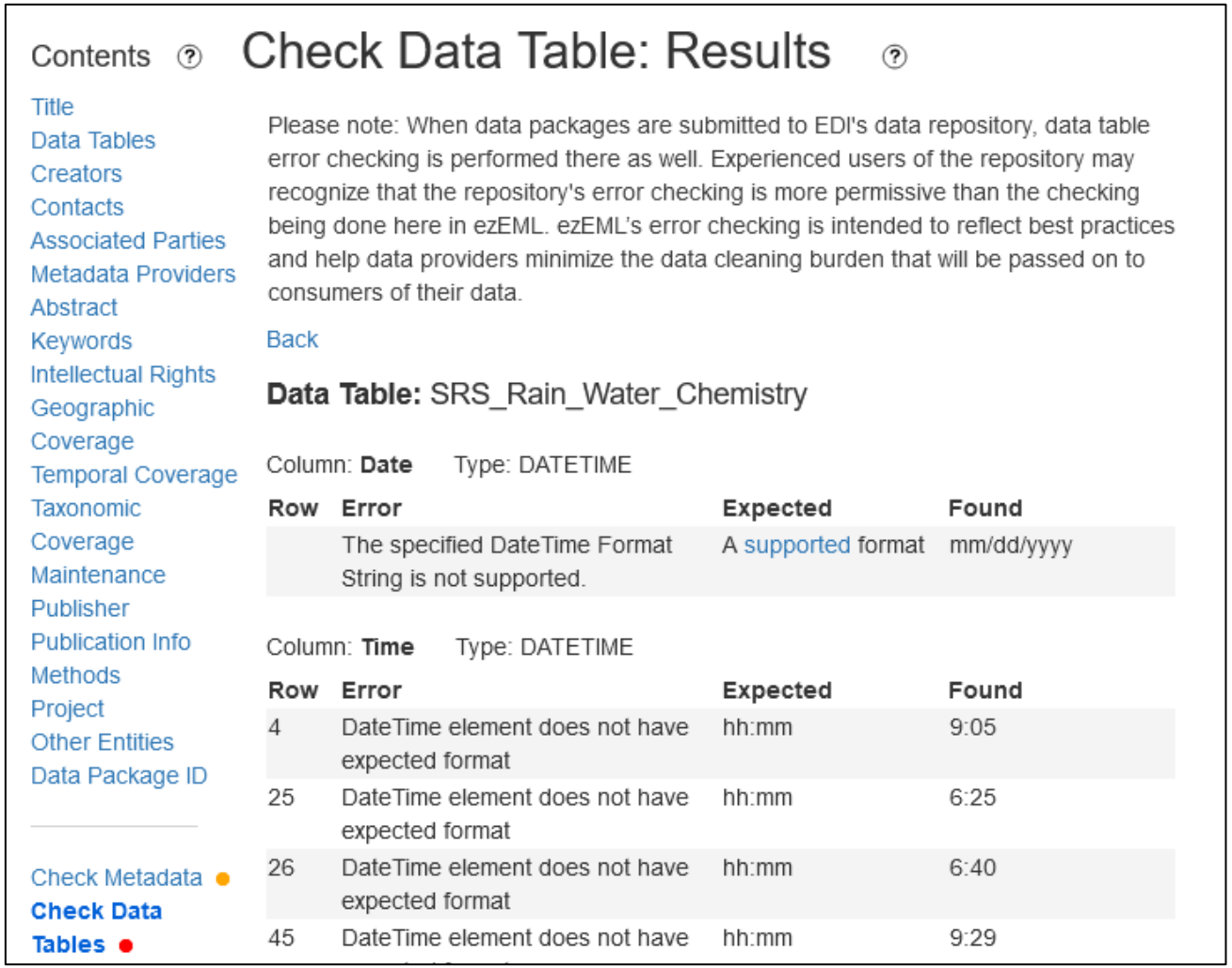
ezEML's Role in Data Publishing Life Cycle

- Can create metadata for tables and other data entities (e.g. model code, imagery, and other nontabular data files)



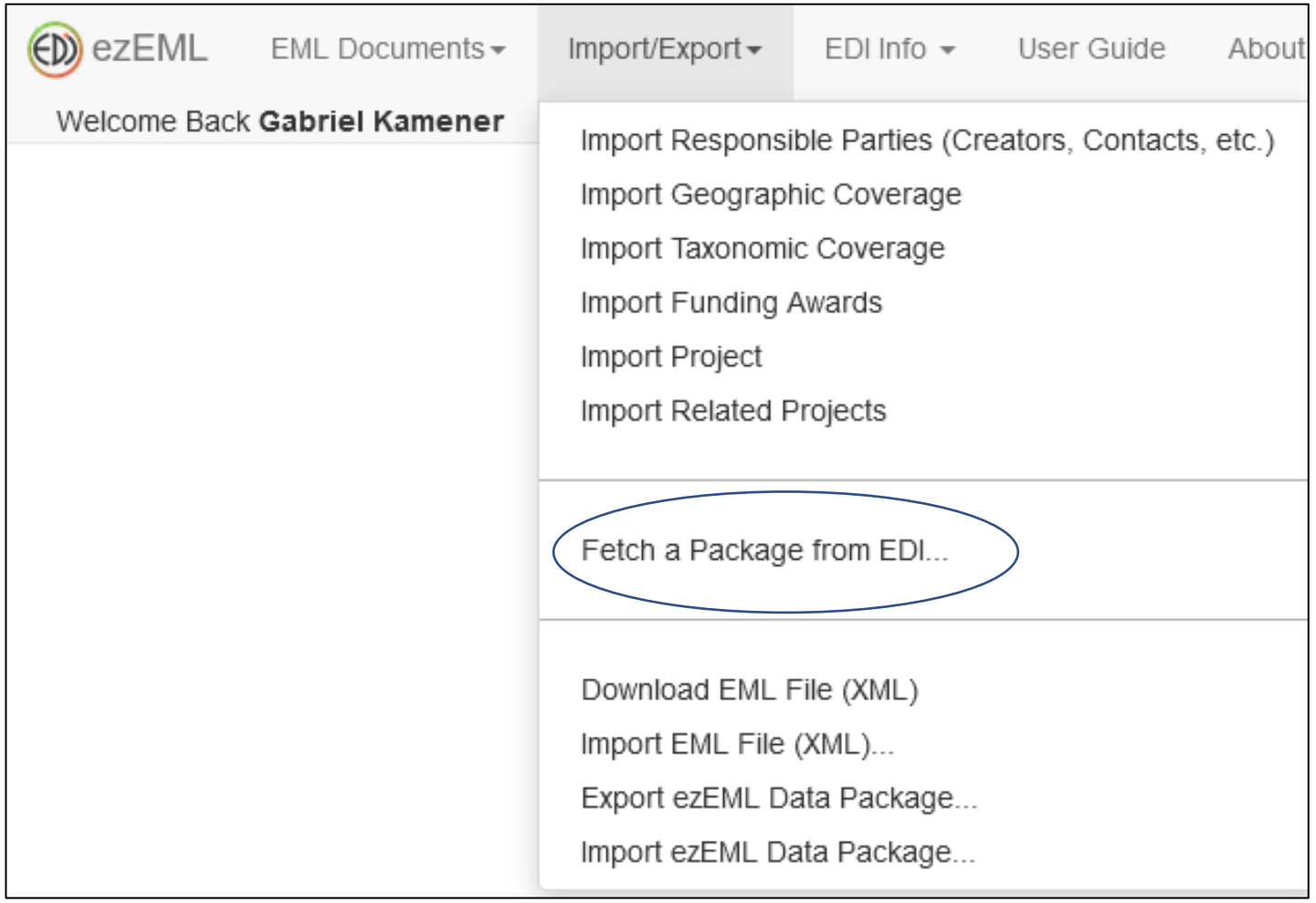
ezEML's Role in Data Publishing Life Cycle

- Quality checks data and metadata
- Provides feedback on potential showstoppers (errors) and points to investigate (warnings)



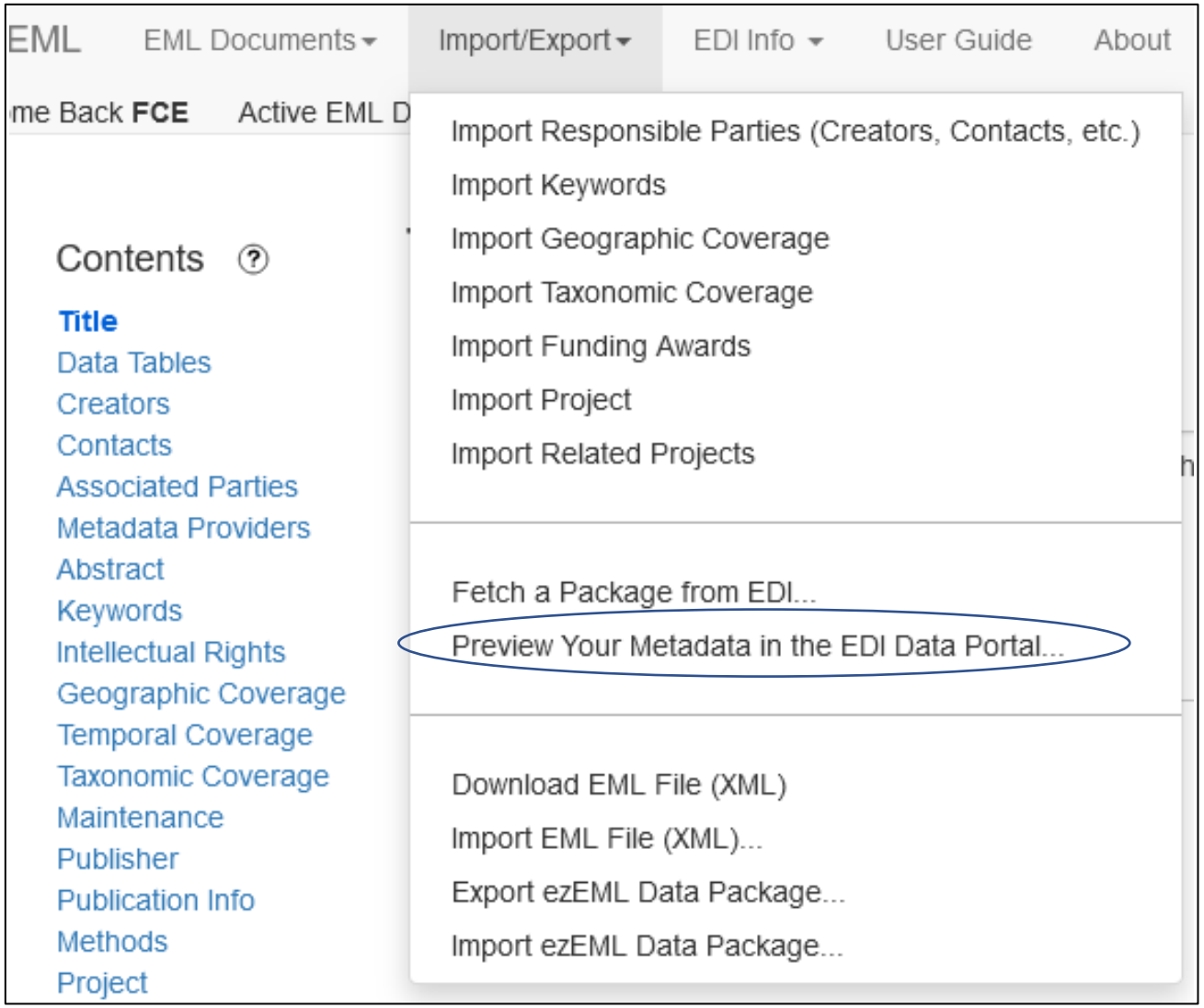
ezEML's Role in Data Publishing Life Cycle

- Can also "fetch" existing packages from EDI to facilitate updates



ezEML's Role in Data Publishing Life Cycle

- Preview how your package will look on the EDI Data Portal

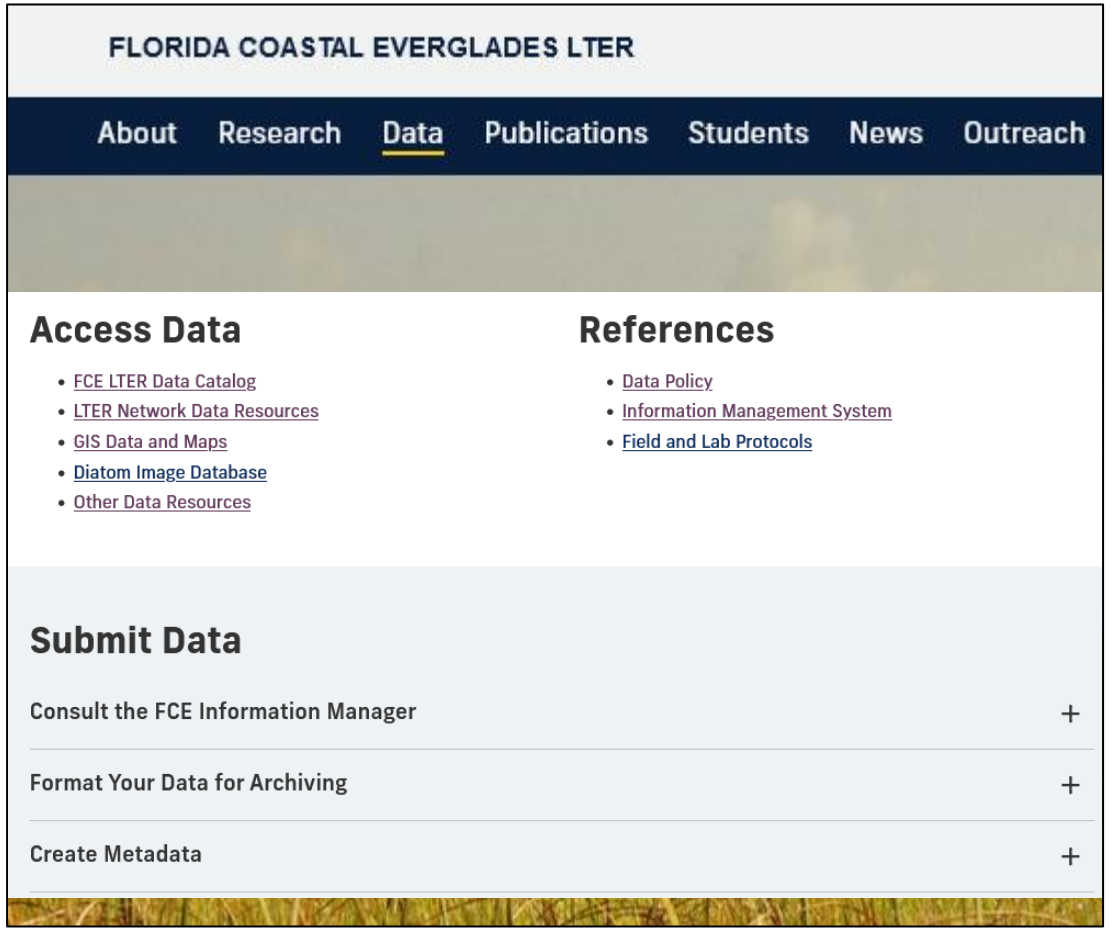


ezEML's Role in Data Publishing Life Cycle

- Preview how your package will look on the EDI Data Portal

Submitting Your Data to the FCE IM

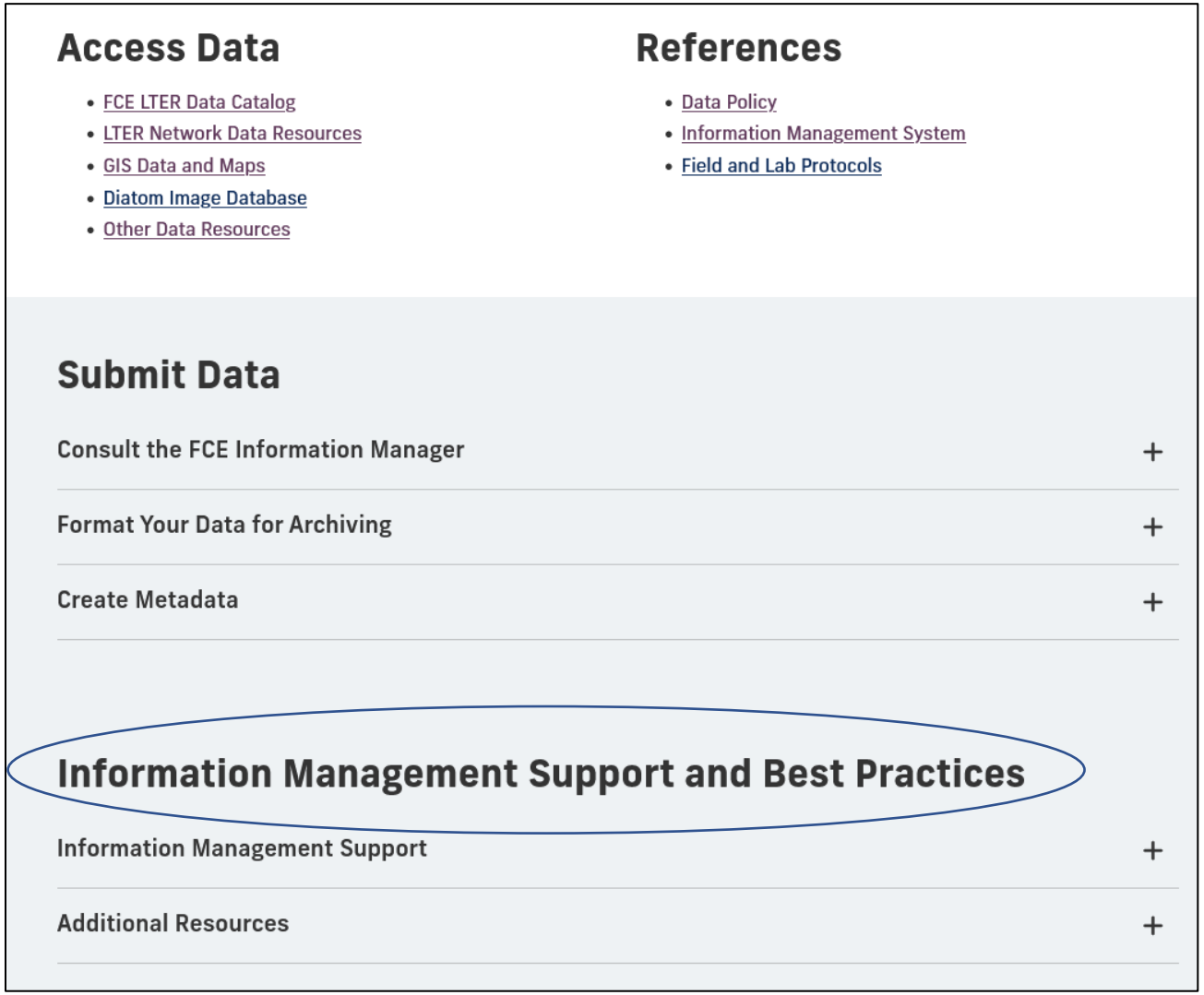
1. Review the FCE Data page!
2. Contact the FCE IM to review your data and to receive an FCE dataset ID
3. Accept ezEML collaboration invite from FCE IM
4. Enter data and metadata into ezEML
5. Review with FCE IM



https://fcelter.fiu.edu/data

IM Support at FCE

- Section on FCE Data page!
- Contains IM support information and a growing collection of links to resources on best practices



https://fcelter.fiu.edu/data

IM Support at FCE

- Weekly IM office hours at MMC-CASE 186B, Thursdays 2:00-7:00 pm
- IM support at FCE student Think Tank events
- Available by email or appointment (in-person or Zoom)

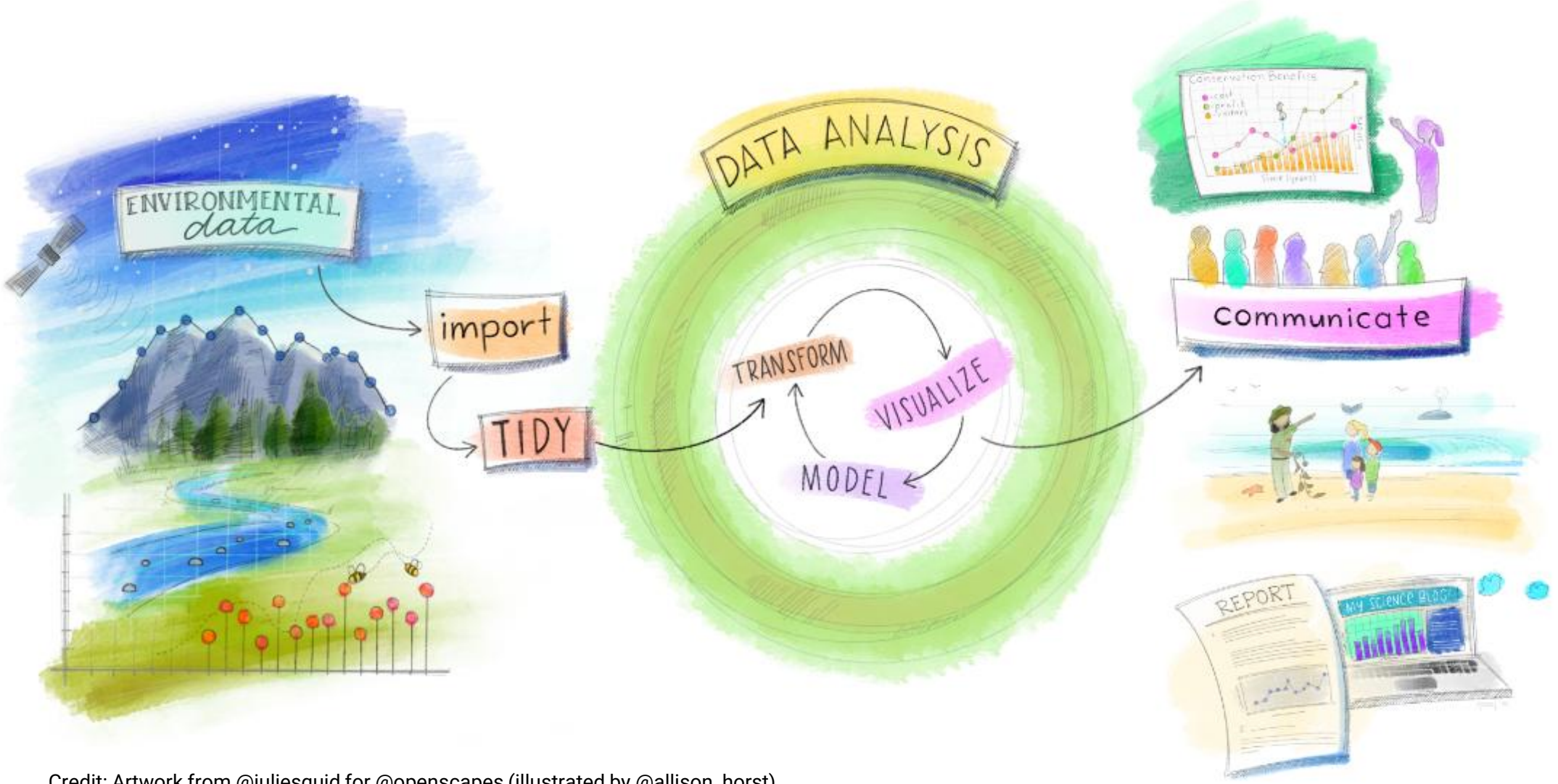
Resources and References

- FCE LTER data page - https://fcelter.fiu.edu/data
- FCE ezEML instructions (PDF file)
- Tips for submitting FCE data and metadata (PDF file)
- FCE GitHub repository containing this presentation and example metadata documents: https://github.com/FCE-LTER/intro_to_fce_data_management_and_publication
- Briney, K. A., Coates, H. L., & Goben, A. (2020). Foundational practices of research data management. Research Ideas and Outcomes 6: e56508. https://doi.org/10.3897/rio.6.e56508
- Briney, K. (2023). The Research Data Management Workbook. Caltech Library. https://doi.org/10.7907/z6czh-7zx60
- Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. The American Statistician, 72(1), 2-10. https://doi.org/10.1080/00031305.2017.1375989

Resources and References

- Casey O'Hara and Li Kui (2025), AI for Ecologists. https://nceas-learning-hub.github.io/ai_for_nceas.
- Cooper, N., Hsing, P.-Y., Almarzouq, B., Baldauf, S., Chatterjee, N., Plomp, E., Strydom, T., Takola, E., & Zagrodzka, Z. (2025). Guide to Reproducible Code (v2.0.1). Zenodo. https://doi.org/10.5281/zenodo.17855982
- Data Carpentry course episode: Introduction to R and RStudio
- Elmendorf (2026), NWT LTER AI Tools Introduction. https://nwtlter.github.io/data_skills/AI_intro_to_AI.html
- Kleindl, P. M., Wachnicka, A., Sah, J. P., Ross, M. S., & Gaiser, E. E. (2025). Hydrology drives facilitative and competitive strategies in freshwater macrophyte and microbial communities. *Ecosphere*, *16*(9), e70384. https://doi.org/10.1002/ecs2.70384

Questions and Discussion



Credit: Artwork from @juliesquid for @openscapes (illustrated by @allison_horst).
https://github.com/allisonhorst/stats-illustrations/tree/master/openscapes?fbclid=IwAR0M_Bksd0Hr5IL4Jla7l8m1azBojU8acN5OmZ4QIGv20xcVbBhbkvpGbIw