Research Data Management:

Maximizing Your Research Impact Through Smart Data Practices

Sonia Santana
Research & Data Repository Manager | FIU Libraries
July 24, 2025

What We'll Cover Today

Why data management matters (spoiler: it's not just about compliance!)

The research data lifecycle made simple

Your new best friend: FIU Research Data Portal

How to make your data work harder for your career

Let's Start With Reality...

The Hidden Costs of Poor Data Management

The "Oops" Moments We've All Had:

🔍 "I know I saved that file somewhere..."

📊 "What did this variable name mean again?"

💾 "Wait, which is the final version?"

🤝 "My collaborator can't open my files"

🗓️ "That analysis from last year? Good luck finding it."

The Real Impact

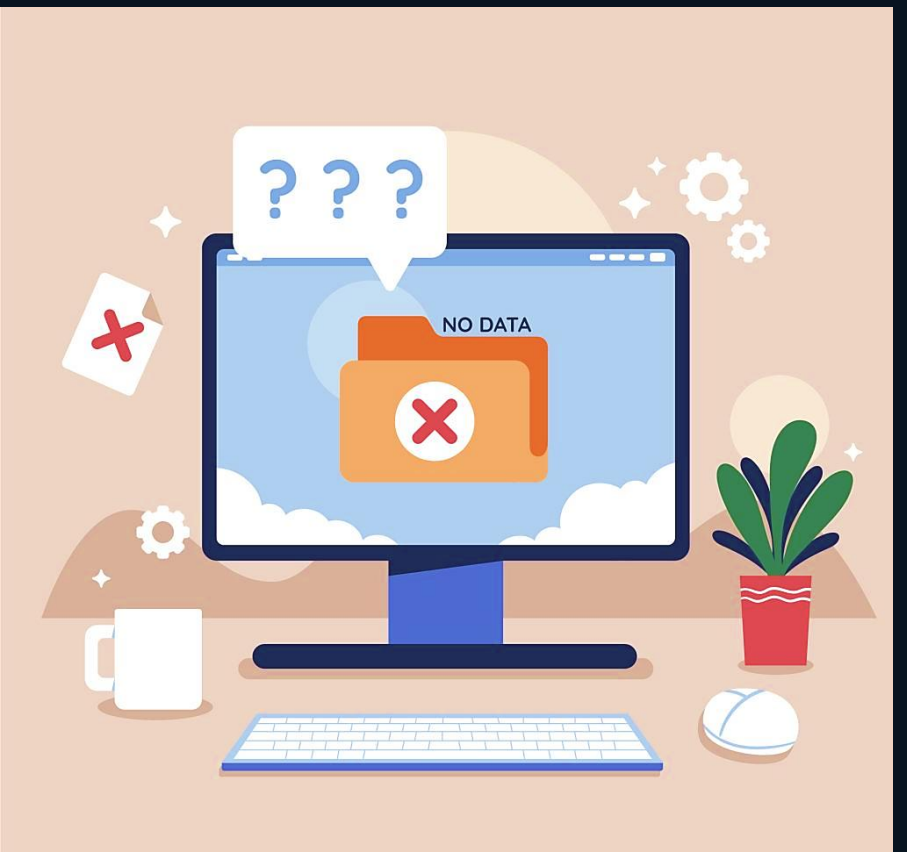
📊 **Research Data Loss Over Time:** Research data faces significant risk of becoming inaccessible over extended periods. Academic studies show that research data availability declines significantly each year after publication, making long-term access increasingly unreliable without proper archiving protocols.

🔄 **Evolving Policy Requirements:** Funding agencies and academic journals are increasingly implementing mandatory data sharing requirements. These policies represent a fundamental shift toward transparency, reproducibility, and public access to research outputs, requiring researchers to adapt their data management practices accordingly.

💰 **Recovery Cost Implications:** Improperly preserved research data can result in enormous financial and time costs for reconstruction, requiring substantial institutional resources that could be better invested in new research.

📧 **"Available Upon Request" Limitations:** This is a common practice in academic publishing where authors indicate data availability through direct contact rather than public repositories. This approach frequently fails to provide actual access, as such requests often go unfulfilled due to various practical barriers including outdated contact information, data loss, or simple non-response.

What Puts Your Research Data at Risk?



💻 **Technology Failures:**

- Hard drive crashes and system failures
- Software becoming obsolete or incompatible
- Cloud storage account closures or access issues

👥 **Human Factors:**

- Staff turnover and knowledge gaps
- Poor documentation and file organization
- Accidental deletion or overwriting

🏢 **Institutional Changes:**

- Lab relocations and equipment changes
- Budget cuts affecting storage systems
- Lack of standardized backup procedures

Understanding these risks is the first step to protect your work.

But First... What IS Research Data?

Traditional Data:

- Numbers
- Spreadsheets
- Databases

Also Considered Data:

- Interview transcripts and audio files
- Survey responses
- Images, videos, 3D models
- Code and algorithms
- Protocols and methodologies
- Lab notebook scans

Research data is the information collected from experiments and observations to support findings. The data should be well-documented to help researchers understand and repeat the study.

What is Research Data Management (RDM)?

More Than Just "Saving Your Files"

RDM is the practice of:

- Planning how you'll handle data before you collect it
- Organizing data so future-you (and others) can understand it
- Storing data safely and accessibly
- Sharing data to maximize impact and meet requirements
- Preserving data for long-term value

Think of it as: *Your research GPS – it helps you know where you're going, how to get there, and how others can follow your path*

Why Should You Care?

The Triple Win: You, Your Research, Your Career

🎯 **For You:**

- Save time (no more hunting for files!)
- Reduce stress (everything has its place)
- Work more efficiently with collaborators

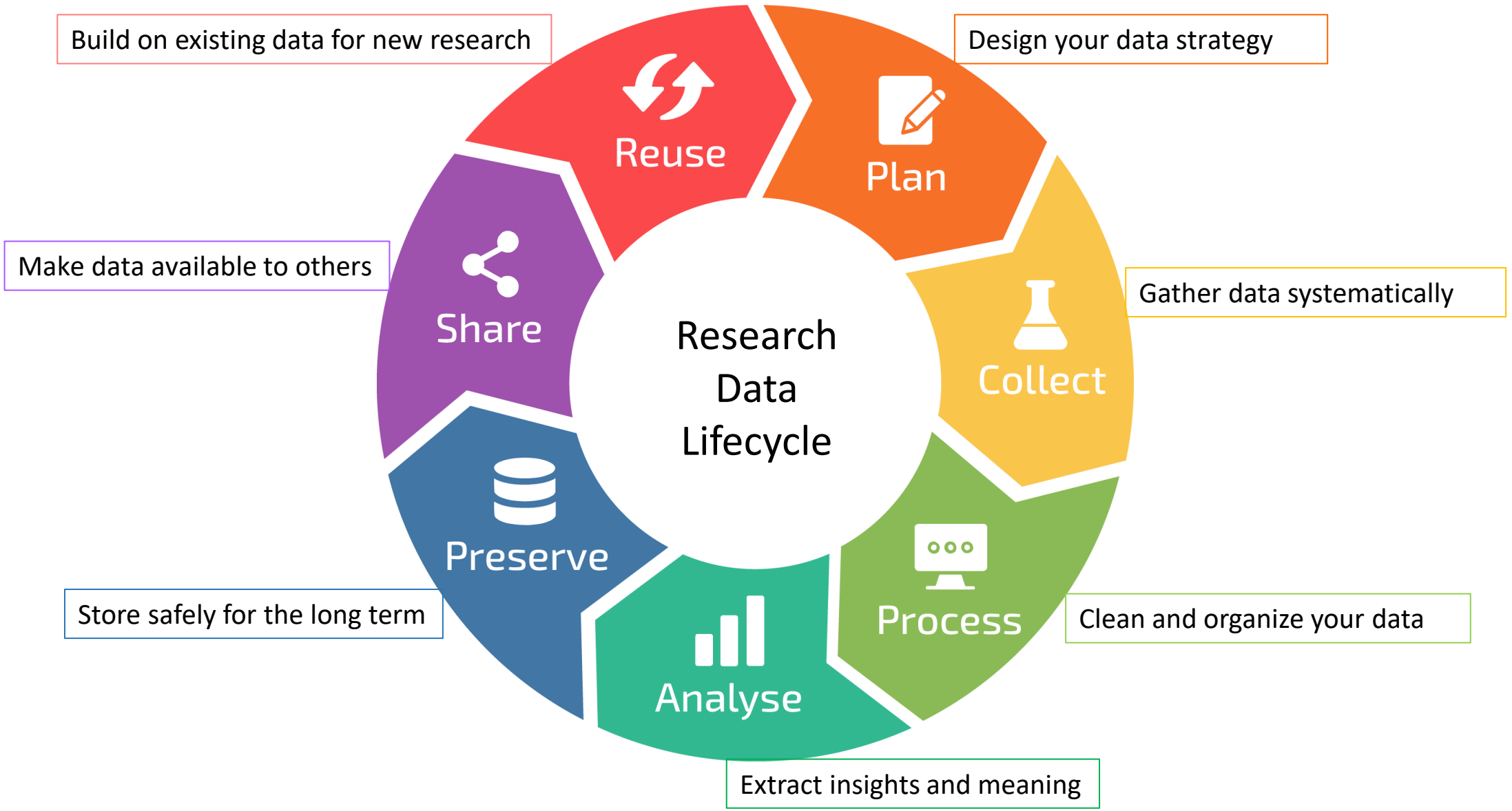
🔬 **For Your Research:**

- Increase reproducibility and credibility
- Enable new discoveries through data reuse
- Meet funder and publisher requirements

🚀 **For Your Career:**

- Boost citation counts through data sharing
- Demonstrate research integrity
- Stand out in grant applications

Good data management isn't extra work – it's working smarter, not harder.



Your Data's Journey from Birth to Legacy

Source: Czech technical university in Prague

Plan & Collect

Setting the Foundation for Success

Key Questions & Considerations:

- What data will I collect and in what format?
- How will I organize and name my files consistently?
- What tools and systems will I use for data collection?
- Who will have access to the data during collection?
- What are the ethical and legal requirements for my data?
- How will I document my data collection process?
- Where will I store my data safely during the project?

Best Practices:

- Create a data management plan before starting
- Establish clear protocols and procedures
- Document everything as you go
- Plan for data security and backup from day one

Process & Analyze

Transforming Raw Data into Insights

Key Questions & Considerations:

- How will I clean and validate my data?
- What analysis methods and tools will I use?
- How will I document my analysis process for reproducibility?
- How will I manage different versions of my datasets?
- What quality control measures will I implement?
- How will I handle missing or problematic data?
- What metadata do I need to capture during analysis?

Best Practices:

- Keep detailed records of all processing steps
- Use version control for both data and code
- Document assumptions and decisions
- Maintain raw data alongside processed versions

Preserve, Share & Reuse

Maximizing Long-term Value and Impact

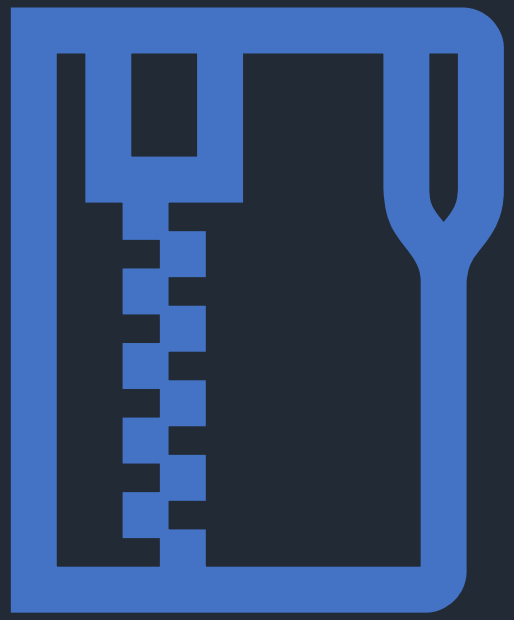
Key Questions & Considerations:

- What data should be preserved for the long term?
- How will I make my data discoverable to others?
- What access restrictions or embargo periods are needed?
- How will I ensure my data remains usable over time?
- What documentation is needed for others to understand my data?
- How can I get credit for sharing my data?
- What licensing terms should I apply?

Best Practices:

- Use standard file formats for preservation
- Create comprehensive documentation and metadata
- Assign persistent identifiers (DOIs)
- Choose appropriate repositories for long-term storage

Data Preparation and Formatting Best Practices

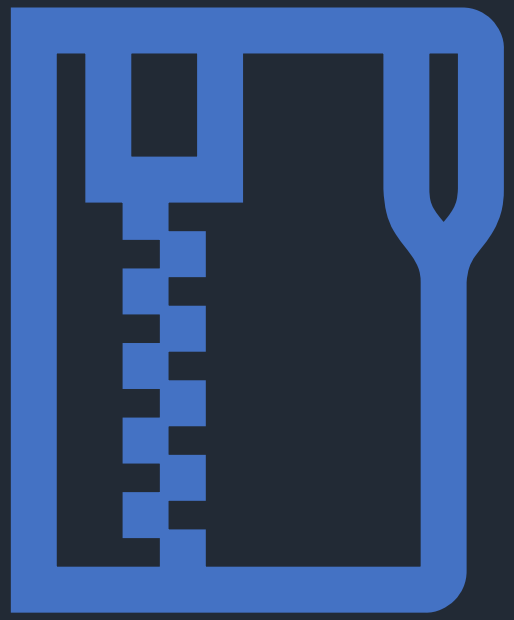


Format Selection and Standards:

Universal Principles:

- **Use open, non-proprietary file formats** (CSV, TXT, PDF/A, JSON)
- **Convert proprietary formats** to preservation-friendly versions when possible
- **Follow discipline-specific standards** when they exist
- **Ensure long-term accessibility** (avoid software-dependent formats)

Data Preparation and Formatting Best Practices



Compression best practices:

- Use ZIP format for collections of related files
- Maintains folder structure and file relationships
- Reduces upload/download time
- Always document compression method used in your README
- Include uncompressed file sizes in documentation

Compression Example:
Instead of uploading 15 separate data files, create: **project_dataset_2025_v01.zip** containing all files with a README explaining the compression method and original file structure.

This approach makes your data more user-friendly while maintaining professional archival standards.

File Naming Conventions - Your Data's First Impression

The Golden Rules for File Names:

🚫 **No Spaces** → Use underscores (_) or hyphens (-)

- `survey_data.csv` **not** `survey data.csv`

🗓️ **ISO Date Format** → YYYY-MM-DD (when dates are needed)

- `2025-07-24` **not** `07-24-2025` or `24-07-2025`

🔢 **Version Control** → Use v01, v02, v03 (leading zeros for proper sorting)

- `analysis_v01.xlsx` **not** `analysis_v1.xlsx`

📝 **Be Descriptive** → Include key identifiers that explain the content

- `temperature_readings.csv` **not** `data.csv`

🔄 **Stay Consistent** → Develop your system and use it across all projects

🌍 **Think Universal** → Avoid personal names, local references, or special characters

- Avoid: `/ \ : * ? " < > | # % & { } $ ! @ + = [ ] ; ,`
- Safe characters: letters, numbers, hyphens (-), underscores (_), periods (.)

File Naming Conventions - Your Data's First Impression

Recommended Formula: Project_DataType_[Date/Version]_Description

✅ **EXCELLENT Examples:**

`fiu_climate_temperature_2025-07-10_raw.csv`
`miami_survey_responses_v03_cleaned.xlsx`
`lab_experiment_protein_v02_results.txt`
`hurricane_interviews_participant-05_transcript.docx`

❌ **AVOID These Mistakes:**

`Final data.xlsx` (spaces, vague)
`data_final_FINAL_v2.csv` (redundant versions)
`temp123.xlsx` (meaningless names)
`María's Data.xlsx` (special characters, spaces)

Pro Tips:

- **Date vs Version**: Use dates for data collection timing, and versions for processing stages
- **Sequential Numbers**: Use padding → `participant-01`, `sample-001`
- **Document Your System**: Create a naming convention guide for your project team to ensure consistency and help new members understand your file organization: Example: 'always use YYYY-MM-DD for dates' or 'use underscores not spaces.'

Folder Structure - Building Your Data Architecture

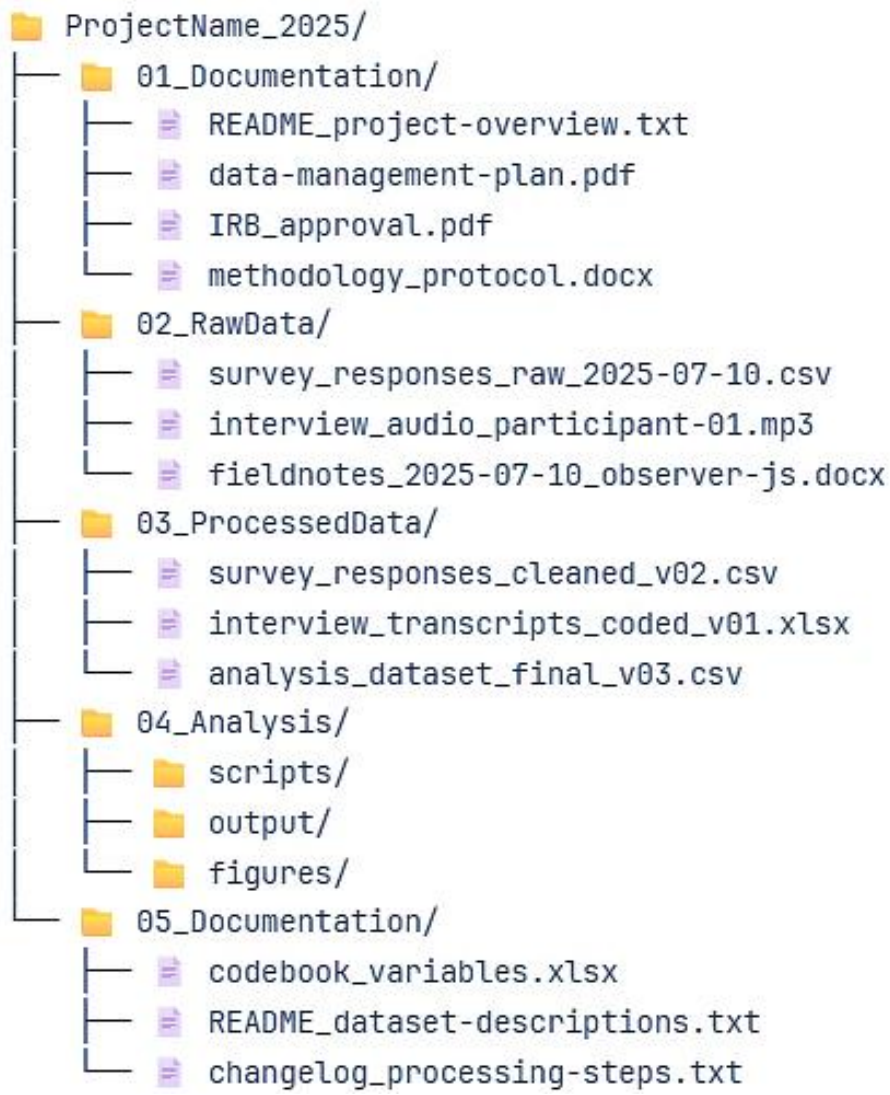
Folder Structure:

- Create logical hierarchies (Project → Data Type → Date)
- Use standard folder names across projects
- Separate raw data from processed data

Folder Naming Best Practices:

- Use numbers for ordering (01_, 02_, 03_)
- Be descriptive but concise
- Separate raw from processed data
- Establish consistent folder organization patterns that you replicate from project to project
- Document your structure in a README file

The Golden Rule: *Organize your data as if someone else (including future you) needs to understand it without your help!*



Version Control - Track Your Data's Evolution

When to Create New Versions:

✅ Major cleaning or processing steps
✅ Before sharing with collaborators
✅ After incorporating feedback
✅ Before publication submission

Smart Version Naming:

dataset_raw_v01.csv ← Original data

dataset_cleaned_v02.csv ← After cleaning

dataset_analysis-ready_v03.csv ← Ready for analysis

dataset_final_v04.csv ← Publication version

The Golden Rules:

🔢 **Use consistent numbering:** v01, v02, v03 (leading zeros!)
📁 **Keep master copies separate** from working copies
🚫 **Never delete previous versions** during active work
📝 **Document what changed** in each version

README Template: Part 1 - Study Overview

Header Information

- **Dataset Title:** [Descriptive name of your study/dataset]
  Created: YYYY-MM-DD | **Contact:** your.email@institution.edu

What This Data Contains

- Brief description of research purpose and main findings

Data Collection

- **When:** Collection period (start - end dates)
- **How:** Methods used (survey, experiment, observation, etc.)
- **Who:** Sample size and key characteristics
- **Where:** Geographic scope or setting

Documentation - Your Data's User Manual

README Template: Part 2 - Data Structure

Files Included

- filename1.csv - Raw data description (N=xxx)
- filename2.xlsx - Processed/cleaned dataset
- codebook.xlsx - Variable definitions and coding schemes

Key Variables

- participant_id: Unique identifier format
- main_outcome: Primary measure (scale/range)
- key_predictor: Important variables for analysis

demographics: Age, gender, etc.

How to Cite

- [Author, Year. Dataset Title. Repository. DOI/URL]
- Include usage restrictions or data access requirements

Documentation - Your Data's User Manual

Study Overview

Data Structure

README.txt
==========

Forest Bird Diversity Survey

Created: 2025-06-01 | **Contact:** j.thompson@forestry.edu
Principal Investigator: Dr. James Thompson, Ecology Department

What This Data Contains

Bird species counts and diversity measurements across old-growth and managed forest sites to assess habitat quality and conservation effectiveness.

Data Collection

- When: April - August 2025 (breeding season)
- How: Point counts and mist net surveys at dawn
- Who: 45 forest plots, 3 observers
- Where: Pacific Northwest, Oregon and Washington

Files Included

- bird_counts.csv - Species abundance data (N=45 sites)
- habitat_data.xlsx - Forest structure measurements
- species_codes.xlsx - Bird species reference list

Key Variables

- site_id: Forest plot identifier (F001-F045)
- species_count: Number of individuals observed
- shannon_index: Species diversity measure
- forest_type: Old-growth vs. managed
- canopy_cover: Percentage canopy coverage

How to Cite

Thompson, J. (2025). Forest Bird Diversity Survey [Data file]. Forestry Research Archive. https://doi.org/10.xxxx/birds.2025.001

==========
Research permits: USFS-2025-042, WDFW-2025-089
Data available for research use only
==========

Metadata Essentials - Making Your Data Discoverable

🔍 **Makes data discoverable in searches**

- Without metadata, your data is essentially invisible to search engines and databases
- Metadata acts like tags or keywords that help others find your dataset when they search for relevant topics

📖 **Provides essential context without requiring file downloads**

- People can read the metadata description to understand what your data contains, how it was collected, and whether it's relevant to their needs
- This saves time and bandwidth - no need to download large files just to see if they're useful

🔄 **Enables informed reuse by explaining data structure and methods**

- Future researchers (including yourself) can understand how to properly use and interpret the data
- Explains what each variable means, how data was collected, what units were used, etc.
- Prevents misinterpretation or misuse of the data

📚 **Supports long-term preservation and accessibility**

- Metadata ensures that even years later, people can still understand and use your data
- Provides the context needed to maintain data integrity over time
- Helps data repositories properly archive and maintain your datasets

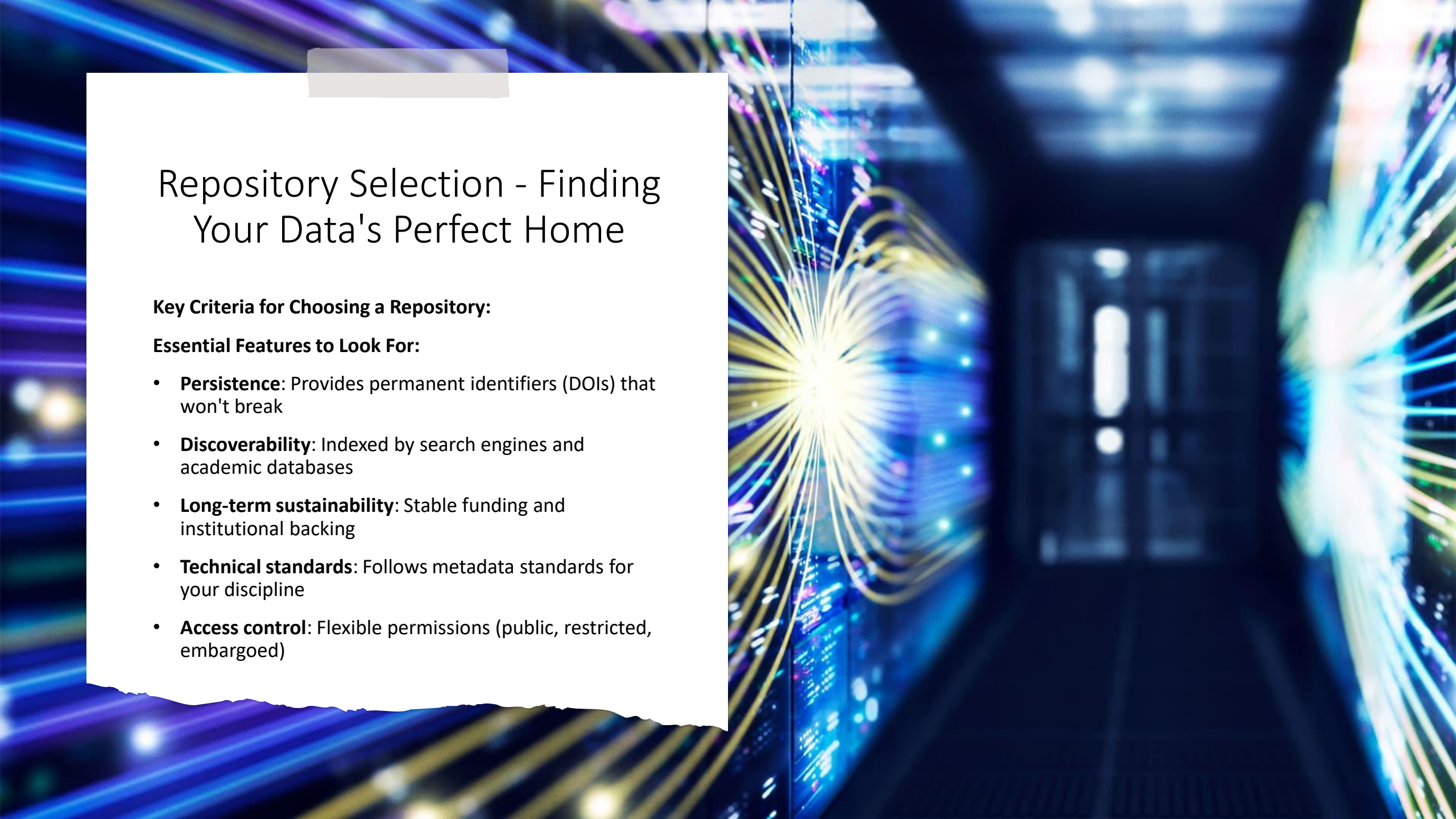
Metadata Essentials - Making Your Data Discoverable

Metadata Quality Levels:

- 🥉 **Basic:** Title + Creator + Description + Date + Subject
- 🥈 **Enhanced:** + Keywords + Coverage + Methodology + Related Publications
- 🥇 **Comprehensive:** + Funding + Variables + Processing Steps + Quality Measures

Target the "Enhanced" level or higher for maximum discoverability and impact!

Remember: Rich metadata transforms invisible data into discoverable, reusable research assets!



Repository Selection - Finding Your Data's Perfect Home

Key Criteria for Choosing a Repository:

Essential Features to Look For:

- **Persistence**: Provides permanent identifiers (DOIs) that won't break
- **Discoverability**: Indexed by search engines and academic databases
- **Long-term sustainability**: Stable funding and institutional backing
- **Technical standards**: Follows metadata standards for your discipline
- **Access control**: Flexible permissions (public, restricted, embargoed)

Repository Selection - Finding Your Data's Perfect Home

Decision Framework:

1. **Check requirements first**: Funder mandates, journal policies
2. **Consider your audience**: Who needs to find and use your data?
3. **Evaluate support**: What help is available for uploading and curation?
4. **Think long-term**: Will this repository exist in 10+ years?

***Remember**: The best repository is one that ensures your data remains accessible, discoverable, and properly credited for the long term.*

How to Find Data Repositories - Your Discovery Toolkit

Start with These Key Resources

Discipline-Specific Recommendations:

- Ask your professional associations and journals
- Check funder requirements and recommended repositories
- Look at where colleagues in your field are sharing data

Publisher and Funder Resources:

- Journal data availability statements often list preferred repositories
- Grant agency websites typically provide repository recommendations
- Check your institution's library research guides

How to Find Data Repositories - Your Discovery Toolkit

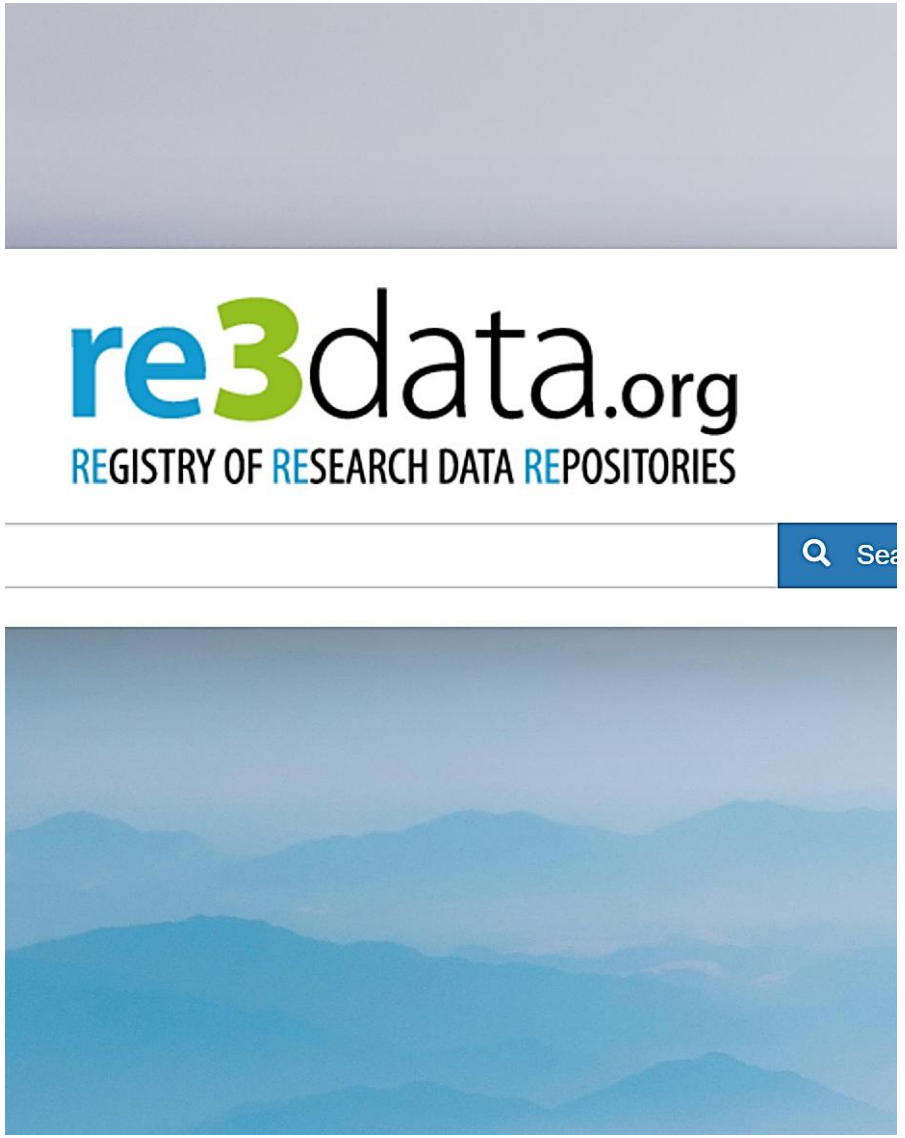
Registry of Research Data Repositories (re3data.org)

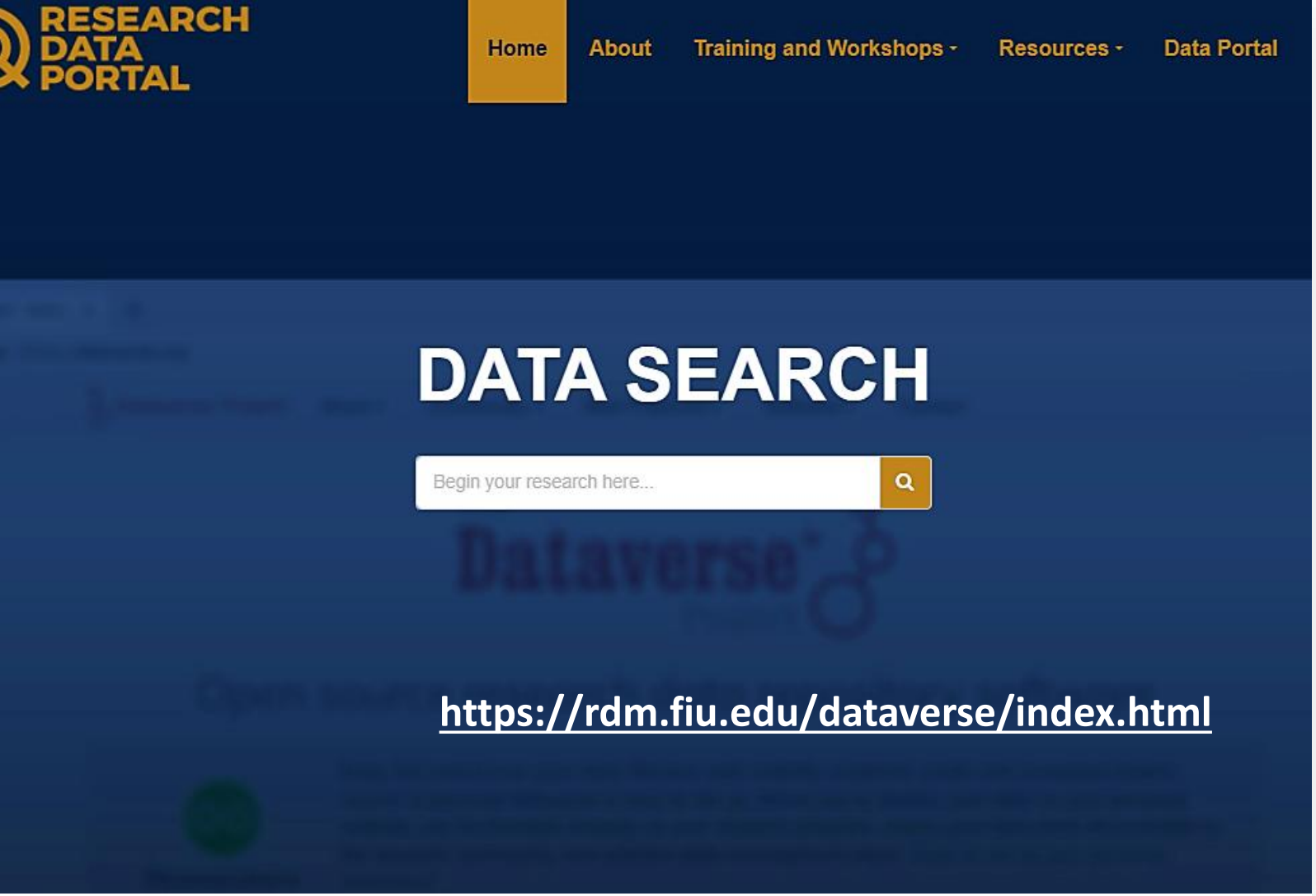
- Comprehensive global registry of 3,000+ repositories
- Filter by subject area, content type, and access conditions
- Provides detailed repository profiles and standards

Pro Tips:

- Start with domain-specific repositories when available
- Consider general repositories (like Zenodo) for interdisciplinary work
- Always verify repository sustainability and standards before committing

Remember: The best repository is one that serves your research community and ensures long-term access!





Meet Your New Research Partner

FIU Research Data Portal

- Built on Dataverse technology (developed by Harvard University)
- Supported by FIU Libraries for the university research community
- Secure, reliable, and professionally maintained

🔧 **What It Does:**

- Stores your research data safely
- Creates DOIs for your datasets (perfect for citations!)
- Provides controlled access sharing
- Ensures long-term preservation
- Integrates with your existing workflow

It's like having a professional librarian for your data, available 24/7

How FIU Research Data Portal Supports You

Source: Author (based on literature)

What Can You Submit?

We Welcome All Types of Research Data

📊 **Data Types We Accept:**

- 📈 **Spreadsheets & Datasets** - Excel, CSV, statistical files
- 📋 **Survey Data** - Questionnaire responses, interview transcripts
- 🗺️ **GIS Data & Imagery** - Maps, satellite data, spatial analysis
- 💻 **Code & Scripts** - Analysis code, software, algorithms
- 📝 **Documentation & Protocols** - Research methods, lab protocols

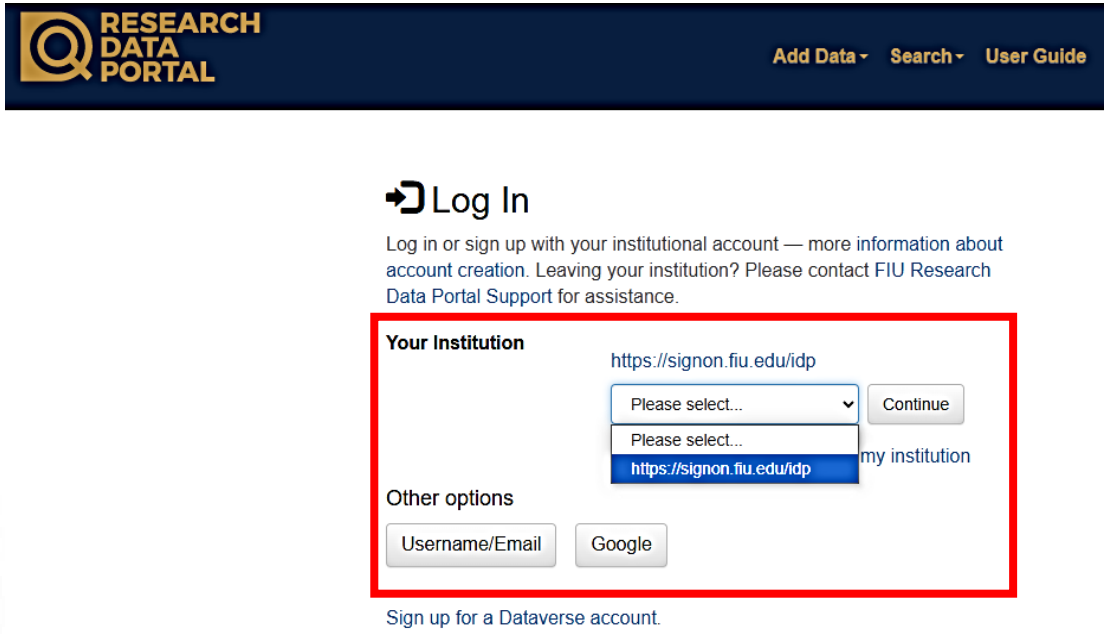
💾 **Storage Details:**

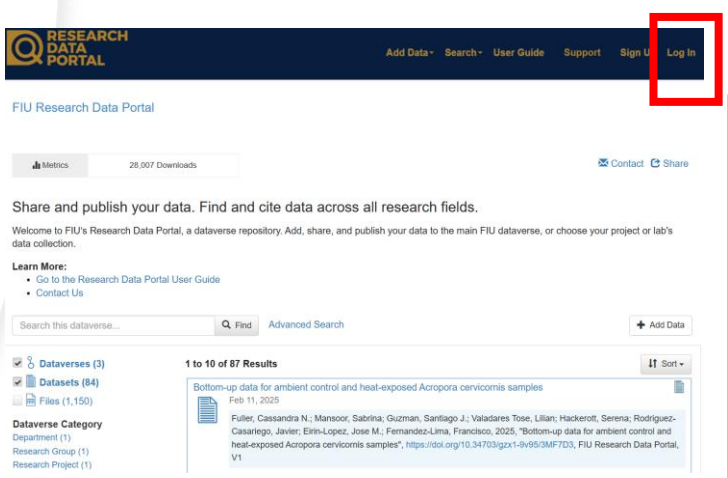
- **Individual files:** Best under 2GB for web upload
- **Total storage:** Up to 100GB per faculty member
- **Large datasets:** We can help with bulk uploads

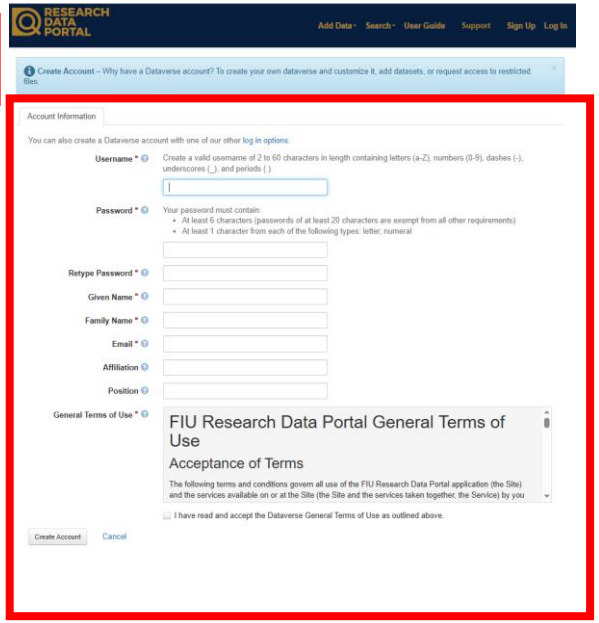
From spreadsheets to code to protocols – if it supports your research, we want to help preserve it!

Login / Creating account

- Access Point: dataverse.fiu.edu
- Click "Login" in top-right navigation
- Login Options:
  - **FIU users**: Select "signon.fiu.edu/idp" from dropdown
  - **Non-FIU users**: Select "Username/Email" option
  - Google account: Select "Google" option
- **First-time non-FIU users** must click "SIGN UP" and complete their registration profile
- **New users** are required to review and accept the Terms of Use







Upload Process Made Simple

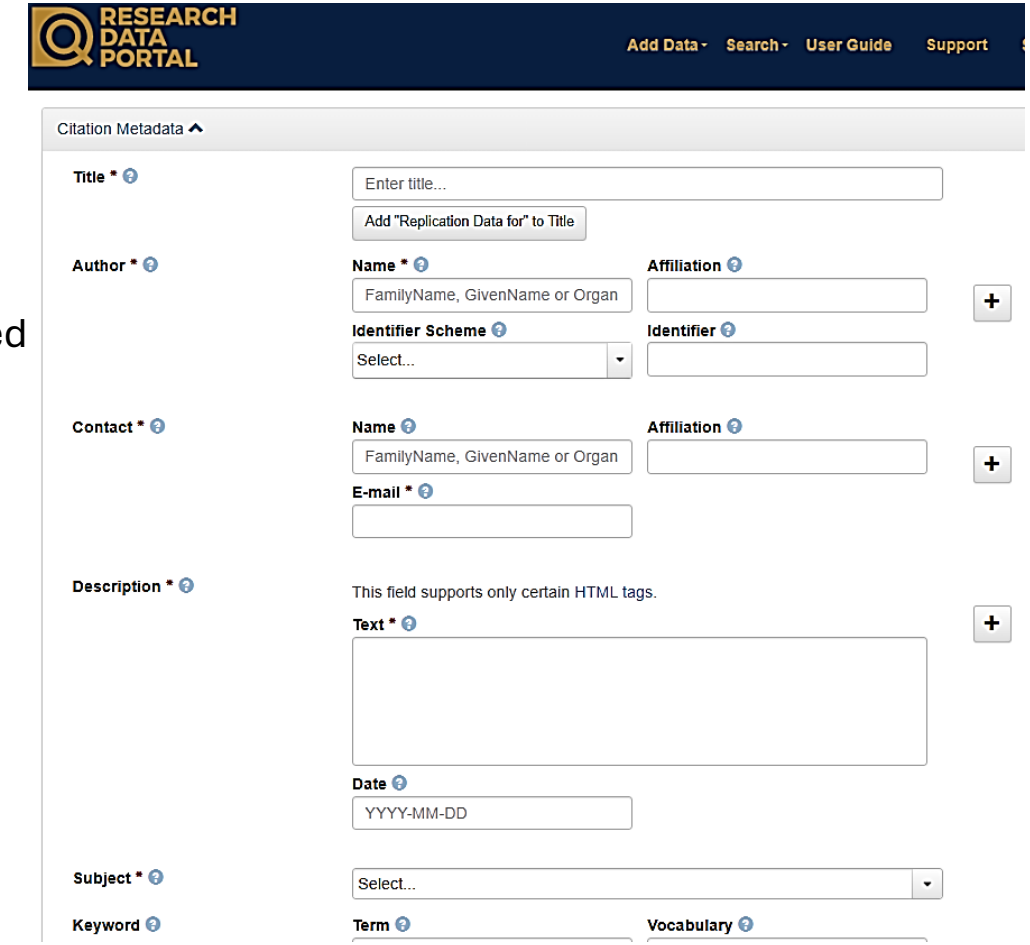
Getting Your Data Into the Portal

✅ **Required Information (The Basics):**

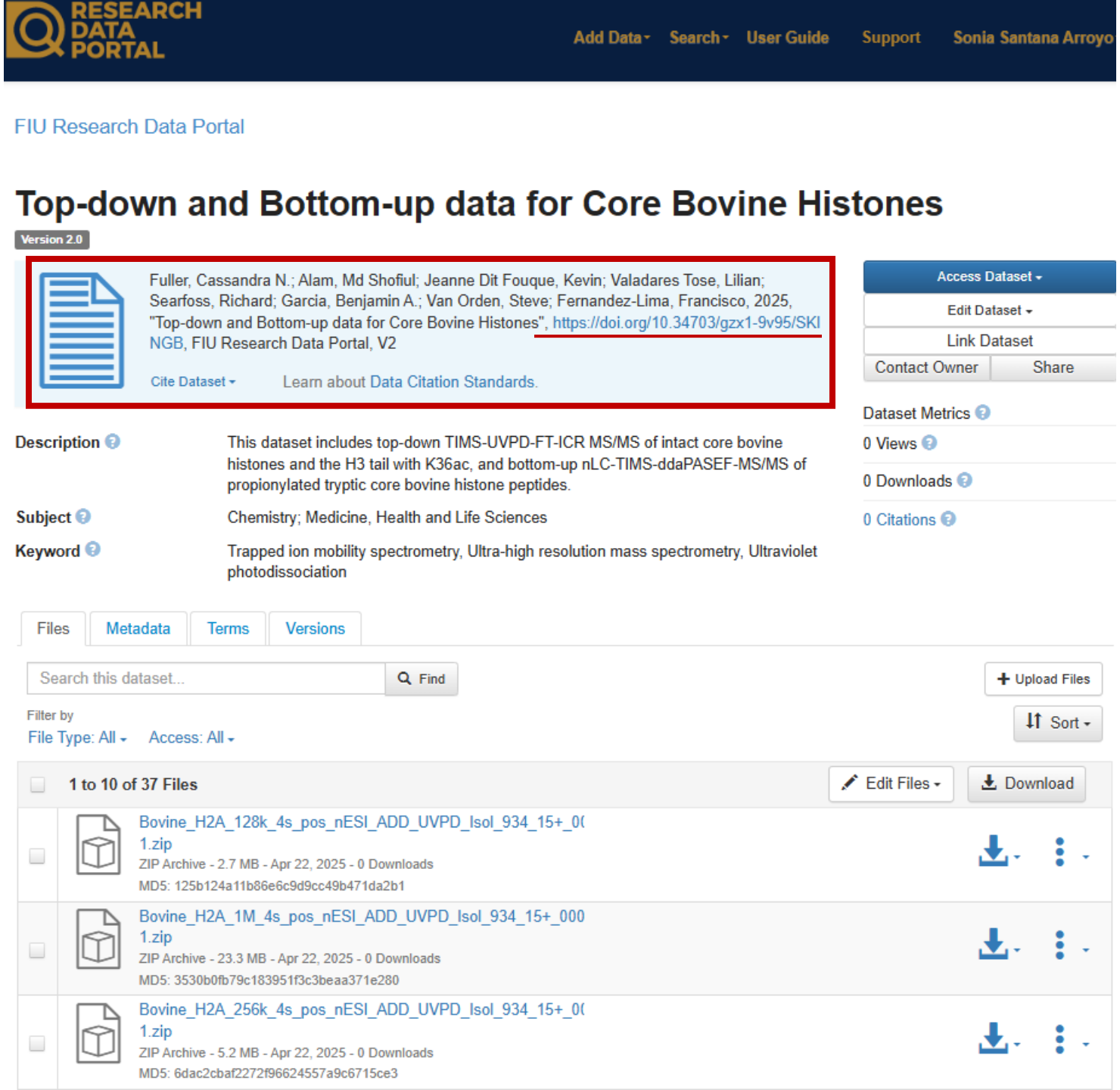
- **Title & Description** - What is this data about?
- **Author & Contact Info -** Who created it and how can they be reached
- **Subject Area** - Which field of research?
- **Keywords** - Help others discover your work

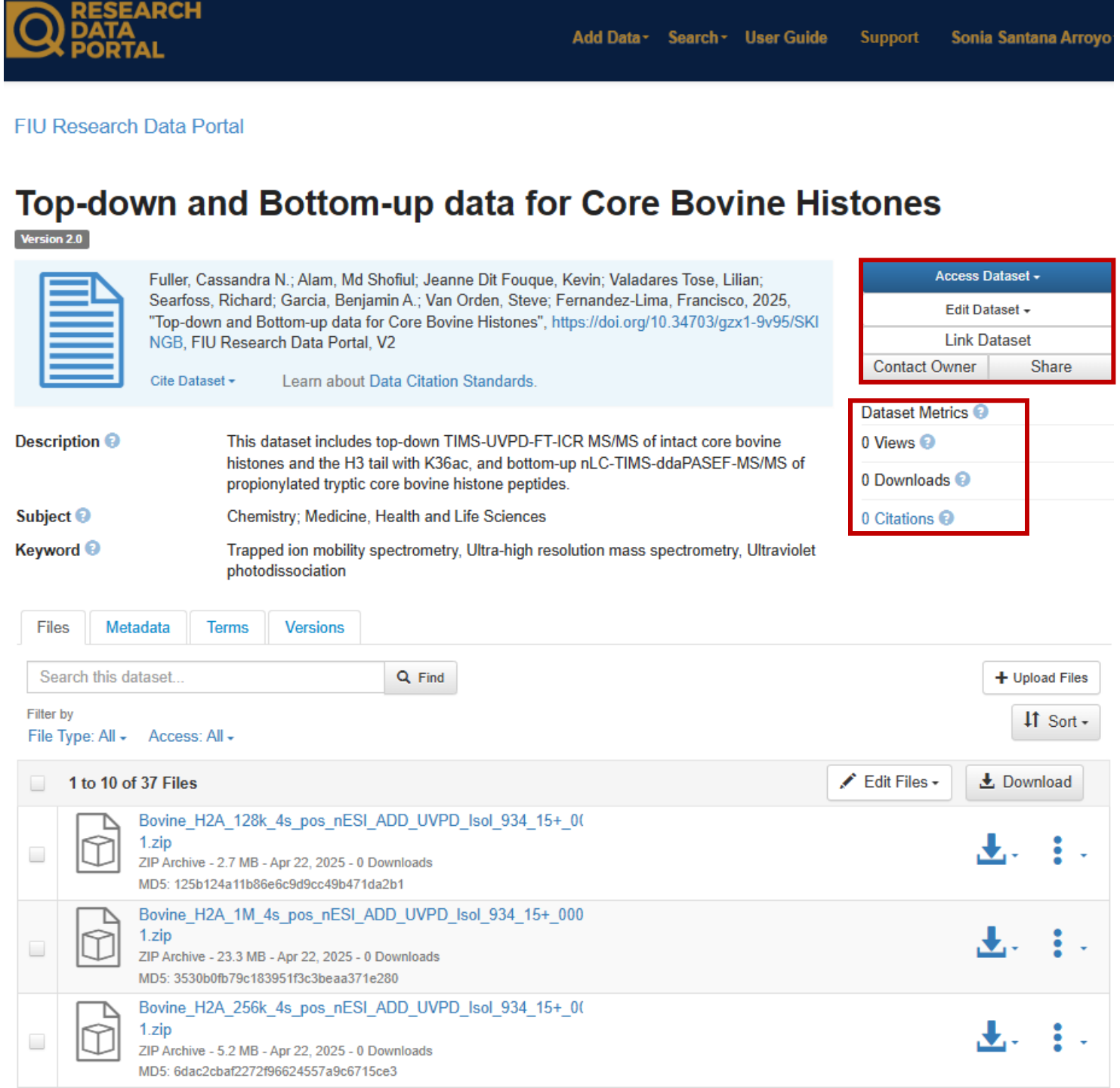
🌟 **Optional Enhancements (Boost Your Impact):**

- **Related Publications** - Link to papers using this data
- **Grant Information** - Connect funding to outputs
- **Notes** - Additional context for users
- **Kind of Data** - Specify data type for better discovery



Persistent Identification:
•DOI assignment for permanent, citable links



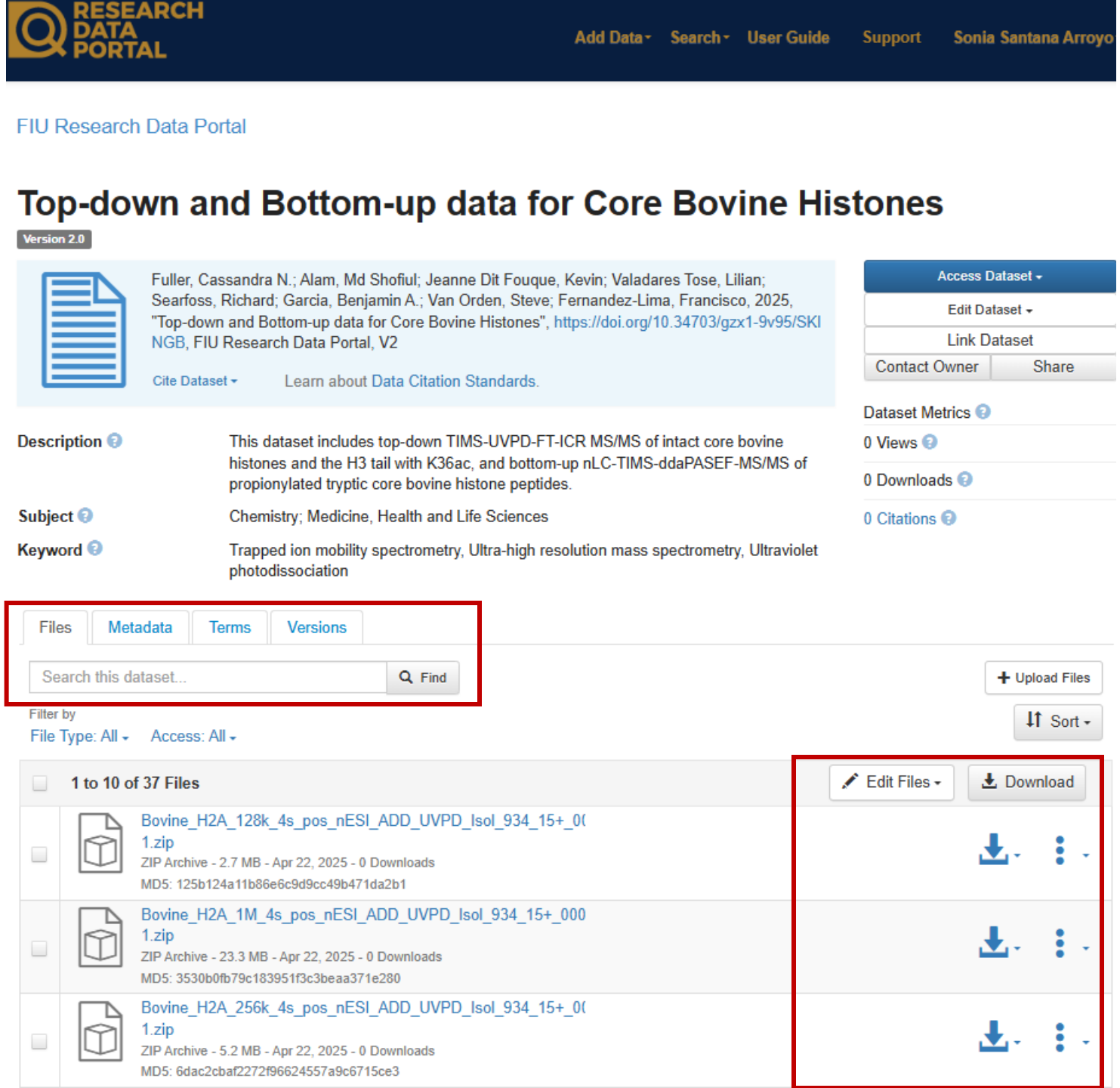


Collaboration Tools:
•Contact owner: "Contact Owner" button
•Sharing: "Share" button
•Dataset management: "Edit Dataset" dropdown, "Link Dataset" option

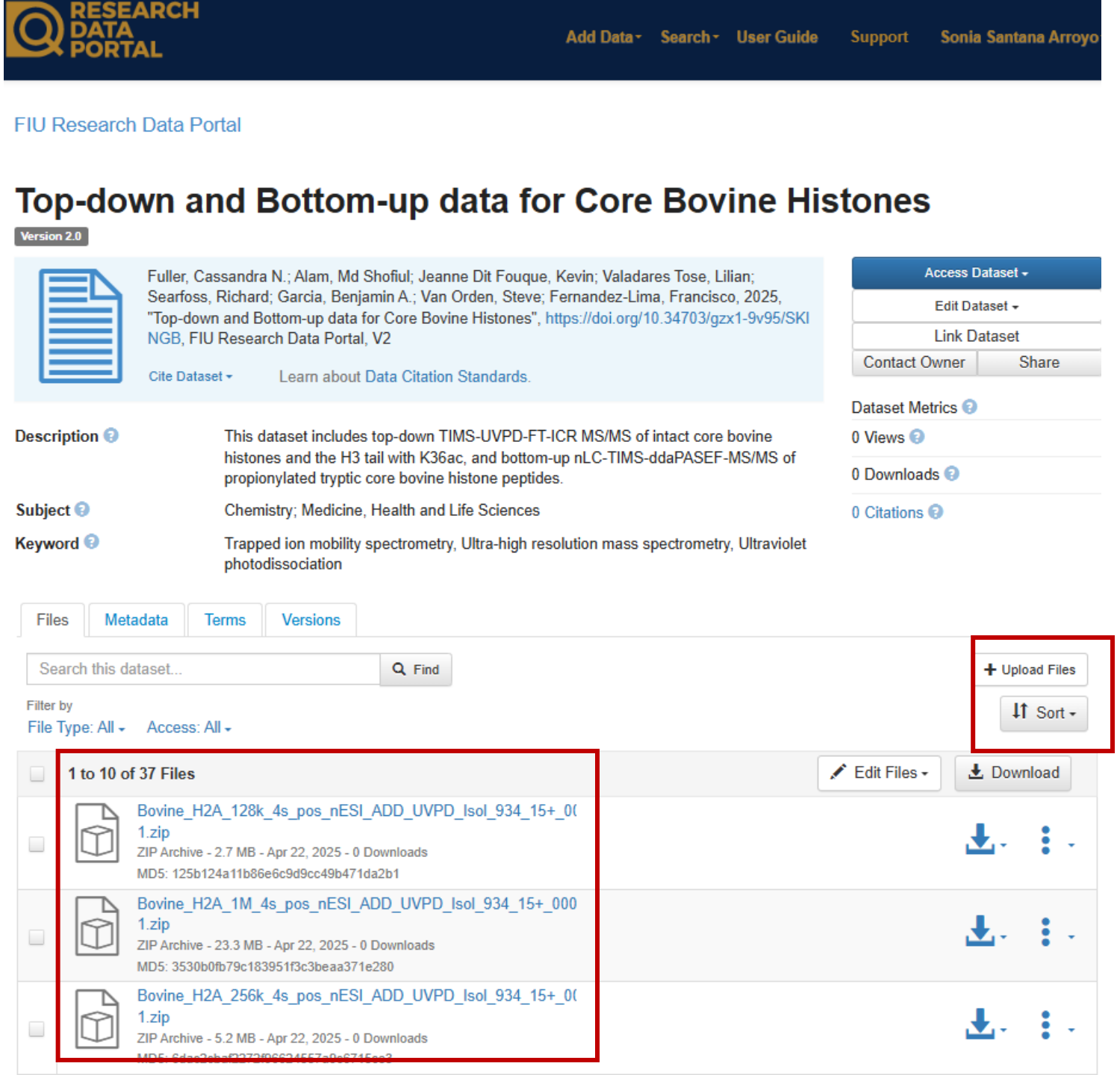
Analytics & Impact Tracking:
•Usage metrics:
•"Dataset Metrics" section showing "Views", "Downloads", "Citations"

User-Friendly Interface:

- Navigation tabs: "Files", "Metadata", "Terms", "Versions" tabs
- **Download options:** Download arrows next to each file and "Download" button
- **File Options menu:** Three-dot menu (⋮) next to each file providing "Edit Options on Metadata", "Restrict", and "Delete" functions



Flexible File Management:
•File listings: Main content area showing zip files with sizes and dates
•Upload capability: "Upload Files" button (top right of file list)
•File organization: "Files" tab



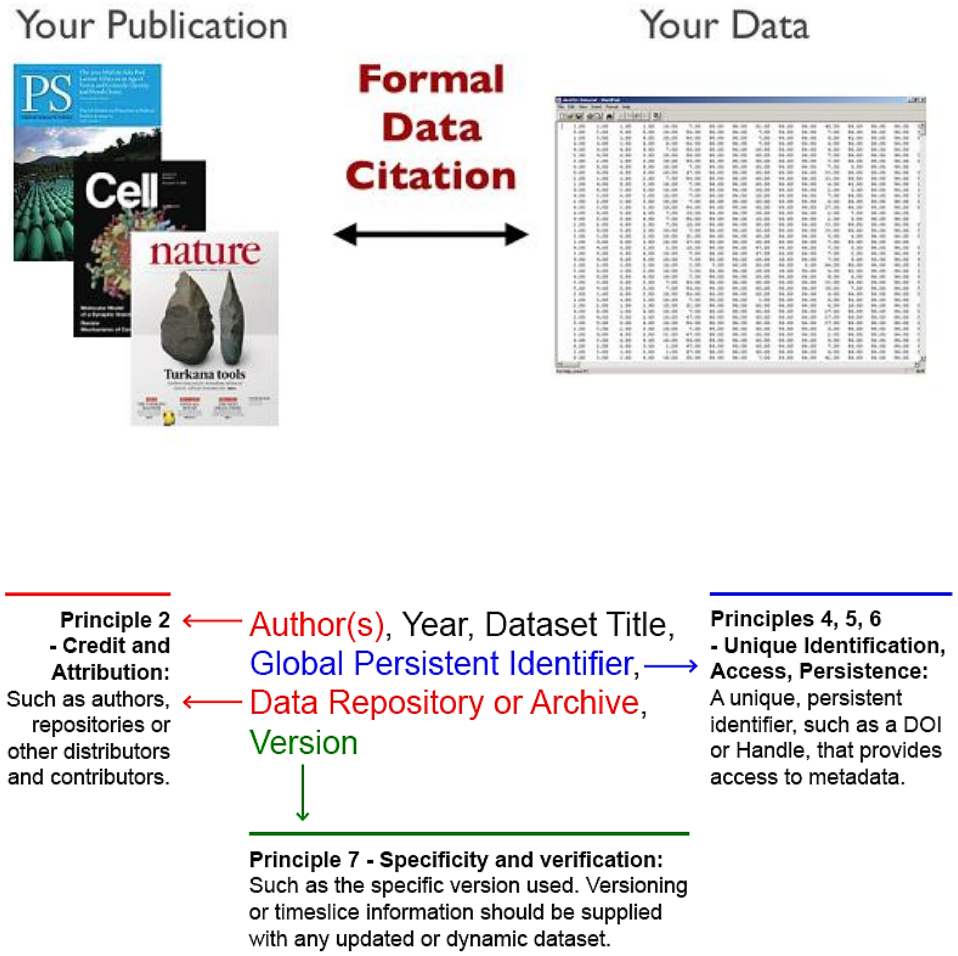
The Citation Advantage

📊 **Make Your Data Work for You:**

- **Permanent DOI Assignment** - Every dataset receives a citable Digital Object Identifier through FIU's Research Dataverse Portal
- **Enhanced Academic Visibility** - Your data becomes discoverable across major research platforms and academic search engines
- **Professional Research Profile** - Link dataset DOIs directly to your ORCID profile to showcase your complete research portfolio
- **Proper Attribution Protection** - Ensure you receive credit when other researchers build upon your work

🎯 **Proven Career Benefits:**

- **25% Citation Boost** - Research shows studies with accessible data receive significantly more citations
- **Direct Dataset Citations** - Researchers can cite your dataset directly using its DOI, giving you credit each time your data is used in other studies
- **Funding Advantage** - Grant agencies increasingly prioritize and favor researchers who demonstrate data sharing practices
- **Long-term Career Investment** - Your datasets continue generating citations and recognition long after initial



Flexible Storage & Organization Options

Your Data, Your Way

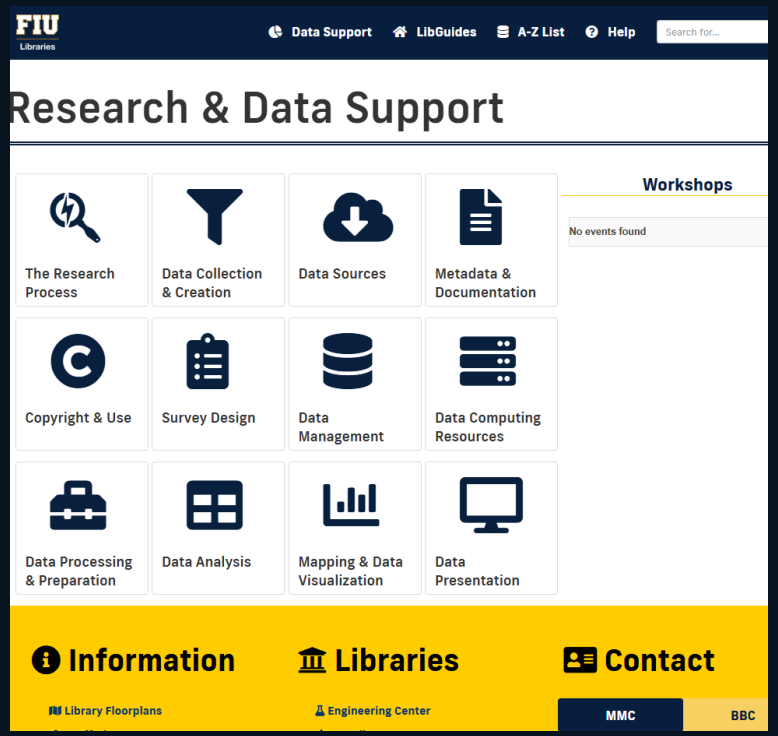
Two Storage Options:

1. **Main Research Data Portal** • Maximum visibility and discoverability • Ideal for data you want widely shared • Appears in global searches and catalogs
2. **Departmental Collections** • Project-specific organization • Perfect for collaborative team projects • Maintains departmental identity and branding

You Stay in Control:

- Set Access Permissions - Public, restricted, or private
- Embargo Sensitive Data - Delay public access as needed
- Update and Version - Keep datasets current
- Track Usage Statistics - See who's using your data

Expert Guidance & Resources



FIU Libraries Research Data Management Team

The FIU Libraries Research Data Management Team serves as an **advisory and consultation resource** for the university's research community. This specialized team provides expert counsel and guidance throughout the data lifecycle, offering researchers strategic direction and best practice recommendations.

We Provide:

- Research data guidance
- Data Management Plan (DMP) assistance
- Metadata creation guidance
- Copyright and licensing advice
- Quality control for metadata of uploaded datasets
- These consultative services are supported by comprehensive resources including detailed LibGuides, standardized templates, and educational workshop series designed to empower researchers with the knowledge and tools needed for effective data management.

Common Concerns Addressed

Let's Tackle the Real Questions

"I don't have time for this"

- Reality: RDM saves time in the long run
- Start small: 15 minutes of organization saves hours later

"My data isn't interesting to others"

- Reality: You can't predict future research needs
- Example: Weather data from the 1800s is now crucial for climate research

"What if someone scoops my research?"

- Reality: You control access and timing
- Options: Embargo periods, restricted access, metadata-only sharing

"I don't know how to organize data properly"

- Reality: We're here to help!
- Support: Training, templates, one-on-one consultations

Getting Started is Easier Than You Think

Your First Steps to Data Management Success

This Week:

1. Audit your current data: What do you have and where is it?
2. Visit the FIU Research Data Portal: https://rdm.fiu.edu/
3. Start with one small dataset: Practice makes perfect

This Month:

1. Create a data management plan for your next project
2. Organize one existing project using best practices
3. Upload your first dataset to the portal

This Semester:

1. Integrate data management into your regular workflow
2. Share at least one dataset publicly
3. Track your citation impact from shared data

Remember: You don't have to be perfect from day one – just better than yesterday!

Key Takeaways

Your Path to Research Data Success

🎯 **Start Small, Think Big:** Begin with one dataset and build your skills progressively

📊 **Organization = Liberation:** Consistent naming and structure free you from data chaos

🔄 **Lifecycle Thinking:** Plan for the entire journey from collection to reuse

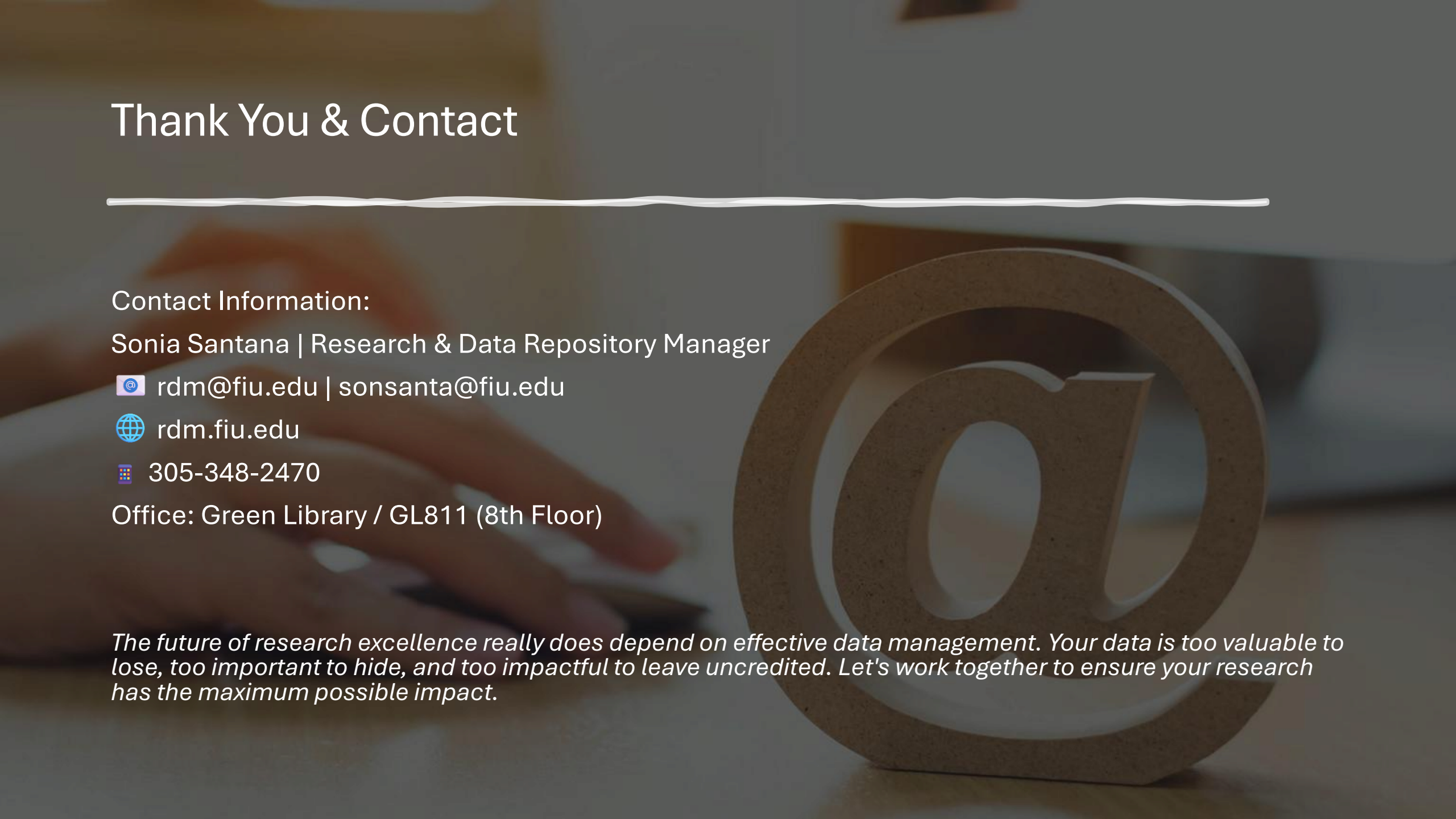
🏛️ **Use Your Resources:** FIU's infrastructure and expertise are here to support you

📈 **Data = Citations:** Shared data amplifies your research impact and career growth

🤝 **Community Matters:** Join the growing movement of researchers committed to open, reproducible science

Good data management isn't just about compliance—it's about doing better research and advancing your career.

Thank You & Contact



Contact Information:

Sonia Santana | Research & Data Repository Manager

📧 rdm@fiu.edu | sonsanta@fiu.edu

🌐 rdm.fiu.edu

📱 305-348-2470

Office: Green Library / GL811 (8th Floor)

The future of research excellence really does depend on effective data management. Your data is too valuable to lose, too important to hide, and too impactful to leave uncredited. Let's work together to ensure your research has the maximum possible impact.

References & Further Reading

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. PloS one, 15(4), e0230416. https://doi.org/10.1371/journal.pone.0230416

Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. *Journal of Documentation*, *75*(6), 1432-1462. https://www.emerald.com/insight/content/doi/10.1108/jd-12-2018-0211/full/html

Fu, Z. J., Juhasz, L., Krefft, J., Ter-Ghazaryan, D., & Guan, B. (2022). *Library-Based Data Curation, Management and Interdisciplinary Research at Florida International University: Reciprocal Use of Data through Collaboration*. https://digitalcommons.fiu.edu/glworks/123/

Gibney, E., & Van Noorden, R. (2013). Scientists losing data at a rapid rate. *Nature*, 10. https://www.nature.com/articles/nature.2013.14416.pdf

Harvard Medical School. (2025). *Data Lifecycle*. https://datamanagement.hms.harvard.edu/plan-design/biomedical-data-lifecycle

National Library of Medicine. (2022). *Research Lifecycle*. https://www.nnlm.gov/guides/data-glossary/research-lifecycle#:~:text=The%20research%20data%20lifecycle%20includes%20everything%20from,institution%2C%20field%20of%20research%2C%20or%20model%20author.

Princeton University. (2025). *Research Lifecycle Guide*. https://researchdata.princeton.edu/research-lifecycle-guide

Research Data Portal - FIU. (2019). *How to use the Research Data Portal – FIU*. https://rdm.fiu.edu/dataverse/our-services.html

Specht, A., Bolton, M. P., Kingsford, B., Specht, R. L., & Belbin, L. (2018). A story of data won, data lost and data re-found: the realities of ecological data preservation. *Biodiversity data journal*, (6), e28073. https://doi.org/10.3897/BDJ.6.e28073

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific data*, 8(1), 192. https://doi.org/10.1038/s41597-021-00981-0

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J. S., Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current biology : CB*, 24(1), 94–97. https://doi.org/10.1016/j.cub.2013.11.014