



HIGHLIGHTS

- Establish clear governance, ownership, and data-sharing expectations before data collection begins.
- Adopt simple, consistent standards for file organization, metadata, and documentation across all project participants.
- Design workflows that can withstand personnel turnover through thorough documentation, centralized storage, and reproducible analyses.
- Plan for long-term stewardship from the outset to ensure data remain accessible, interpretable, and useful beyond the life of the project.

LESSONS LEARNED FROM FIVE YEARS OF SILVA21 DATA MANAGEMENT

Amy Wotherspoon, PhD
Silva21 Scientific Coordinator from 2021 - 2026

¹Department of Forest Resources Management, University of British Columbia
(amy.wotherspoon@ubc.ca)

Modern forestry research increasingly relies on large, interdisciplinary collaborations involving academics, federal and provincial researchers, industry partners, Indigenous communities, and highly qualified personnel (HQP). Research projects developed through these collaborations generate enormous amounts of information, ranging from field measurements and climate records to remote sensing products, modeling outputs, and socio-economic data.

While publishing results in scientific journals often receive the most attention, the ability to effectively manage, document, share, and preserve research data can ultimately determine the long-term success of a project. Experiences from Silva21's five-year research network revealed several common challenges and valuable lessons that can help future projects build stronger foundations from the start.



LESSON 1: GOVERNANCE SHOULD COME BEFORE DATA

One of the most common challenges in collaborative research projects is uncertainty surrounding data ownership, access, publication rights, and long-term maintenance. These issues often emerge after data have already been collected, making them more difficult to resolve. Successful projects establish clear **governance frameworks** at the beginning. This means defining who owns data, who can access it, how it can be shared, and what happens after projects finish to help reduce future conflicts and creates a foundation for collaboration.



LESSON 2: CONSISTENCY OUTWEIGHS PERFECTION

Researchers often develop workflows that suit their individual needs, resulting in different file structures, naming conventions, and data formats across a project. While each system may work





independently, integrating information from multiple groups can become time-consuming and frustrating. Projects benefit from adopting **simple, practical standards** that everyone can follow. Consistent folder structures, file naming conventions, and metadata templates significantly improve efficiency and reduce confusion.



LESSON 3: METADATA IS CRITICAL FOR LONG-TERM VALUE

Data without documentation quickly lose value. Researchers often know and understand the meaning of their own variables, units, methods, and assumptions while actively working on a project, but that knowledge is often lost over time. **Metadata** provides the context needed to interpret and reuse data across different users. Effective metadata should explain what was collected, how it was collected, when it was collected, who collected it, and any limitations associated with the dataset. A useful rule of thumb is that a new researcher should be able to understand a dataset without needing to contact the original creator.



LESSON 4: PEOPLE CHANGE, DATA REMAIN

Graduate students, postdoctoral researchers, and research associates often move on to new opportunities and unfortunately, their data often leaves with them. Projects can reduce this risk by **documenting workflows, analytical methods, and decision-making processes** throughout the project. Readme files, data dictionaries, workflow diagrams, and annotated code help ensure continuity and support future users.



LESSON 5: REPRODUCIBILITY SHOULD BE BUILT INTO EVERYDAY WORKFLOWS

Many research analyses involve manual processing steps that are difficult to track and reproduce. This creates challenges when results need to be verified, updated, or expanded, particularly when these additional steps are being done by someone else. **Automated and documented workflows** improve transparency and reduce errors. Version-controlled scripts and reproducible analytical pipelines allow researchers to regenerate results and understand how decisions were made throughout the project.



LESSON 6: CENTRALIZED STORAGE IMPROVES COLLABORATION

Large projects often struggle with data scattered across personal computers, external hard drives, cloud storage services, and institutional servers. This can lead to duplicate files, outdated versions, and uncertainty regarding which dataset is most current. A **centralized repository** that serves as a single source of truth improves accessibility, consistency, and security. Equally important is ensuring that project participants understand and use the system consistently.



LESSON 7: DATA SHARING IS PRIMARILY A PEOPLE ISSUE

While technology can facilitate data sharing, social and organizational factors are often the greatest barriers. Researchers may have concerns about publication opportunities, data ownership, or potential misuse of information. **Building trust** through clear expectations, transparent policies, and appropriate recognition of contributions encourages a more collaborative culture and improves data sharing outcomes.

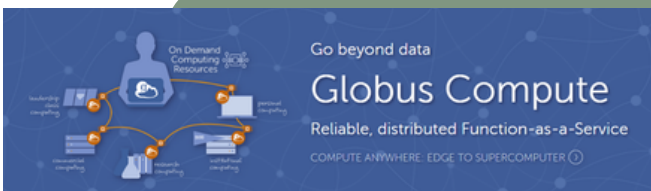


LESSON 8: PLAN FOR THE FUTURE AT THE BEGINNING

Research projects often focus on immediate objectives while overlooking long-term stewardship. When funding ends, questions arise regarding who will maintain repositories, answer questions, and preserve access to project outputs. **Consider sustainability** early to allow projects time to develop archiving plans, identify long-term repositories, and ensure valuable datasets remain accessible to future researchers and decision-makers.

IMPLICATIONS FOR ADAPTIVE SILVCULTURE

Adaptive silviculture depends on the ability to learn from past management actions, monitor outcomes over time, and adjust practices in response to changing environmental conditions. This requires high-quality, well-documented datasets that can be integrated across regions, disciplines, and organizations. Effective data management improves the ability to track treatment responses, evaluate climate adaptation strategies, compare outcomes across sites, and support evidence-based decision making. As forest managers increasingly rely on long-term monitoring, remote sensing, growth and yield models, and climate projections, robust data stewardship becomes a critical component of adaptive silviculture and continuous learning within managed forest ecosystems.



SILVA21 CHOSE GLOBUS COMPUTE AS THEIR CENTRALIZED REPOSITORY.

Here's what we learned:



Great for large file storage and transfer, cross-institutional sharing and accessibility. Desktop tool available for continuous syncing. More users = more storage.



Non-intuitive user interface can make user onboarding difficult and limit greater overall adoption. Requires the right training tools and ongoing permission management.

TIPS FOR USE & MANAGEMENT

- Provide instructional documents, FAQs, and video guides to facilitate the user experience
- Implement required standardize use at the beginning of the project
- Use in conjunction with GitHub to facilitate workflows and automated processing

This document was created from a presentation given by Amy Wotherspoon to the DIVERSE research group in May 2026. To view the original presentation, click the thumbnail.

