Tools for Research Collaboration Transparency and Reproducibility in Research Using R, Git, and GitHub

Research Collaboration & Reproducibility with R

Research increasingly depends on teamwork, shared data, and shared code.

Collaboration platforms support coordination, accountability, and long-term research continuity.

Transparency allows others to clearly understand how research results were produced.

Session Objectives

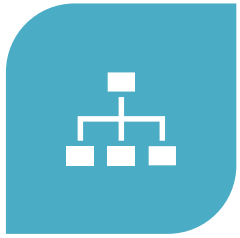
1. Identify major platforms supporting collaboration in modern academic research environments.
2. Understand transparency and reproducibility as core principles of scientific integrity.
3. Learn conceptual foundations of Git and GitHub for research projects.
4. Apply Git and GitHub workflows directly within the R environment.

Why Collaboration in Research Matters

MOST RESEARCH PROJECTS INVOLVE MULTIPLE CONTRIBUTORS ACROSS INSTITUTIONS AND DISCIPLINES.

COLLABORATION IMPROVES EFFICIENCY, QUALITY CONTROL, AND KNOWLEDGE SHARING IN RESEARCH.

SHARED PLATFORMS REDUCE DUPLICATION OF EFFORT AND MISCOMMUNICATION AMONG RESEARCHERS.



CLEAR COLLABORATION STRUCTURES SUPPORT ACCOUNTABILITY AND RESPONSIBILITY AMONG TEAM MEMBERS.

Research Collaboration Platforms

Researchers commonly use email and cloud storage for basic collaboration.

Google Drive and OneDrive allow shared access to documents and data.

Open Science Framework supports structured research project management.

GitHub enables collaborative version control for data and code.

Limitations of Traditional Platforms

Email attachments often create multiple conflicting versions of files.

Cloud storage overwriting causes accidental loss of earlier work.

Tracking individual contributions becomes difficult using traditional platforms.

Analytical decisions are rarely documented clearly in shared folders.

Transparency in Research

Transparency requires open sharing of research methods and data.

Analytical decisions must be clearly documented and justified.

Other researchers should understand how results were produced.

Transparent research builds trust within scientific communities.

Reproducibility in Research

Reproducibility focuses on obtaining identical results using original materials.

Independent researchers should rerun analyses without additional clarification.

Data, code, and documentation must be complete and accessible.

Reproducibility is fundamental to scientific credibility and verification.

Version control tools strongly support reproducible research practices.

Core Reproducibility Question

Scientific claims must be verifiable by independent researchers.

The same data should produce identical results.

The same methods should lead to consistent conclusions.

Missing steps prevent successful reproduction of findings.

Git-based workflows directly address this core question.

What is Git?

Git is a version control system for tracking file changes.

It records every change made to files over time.

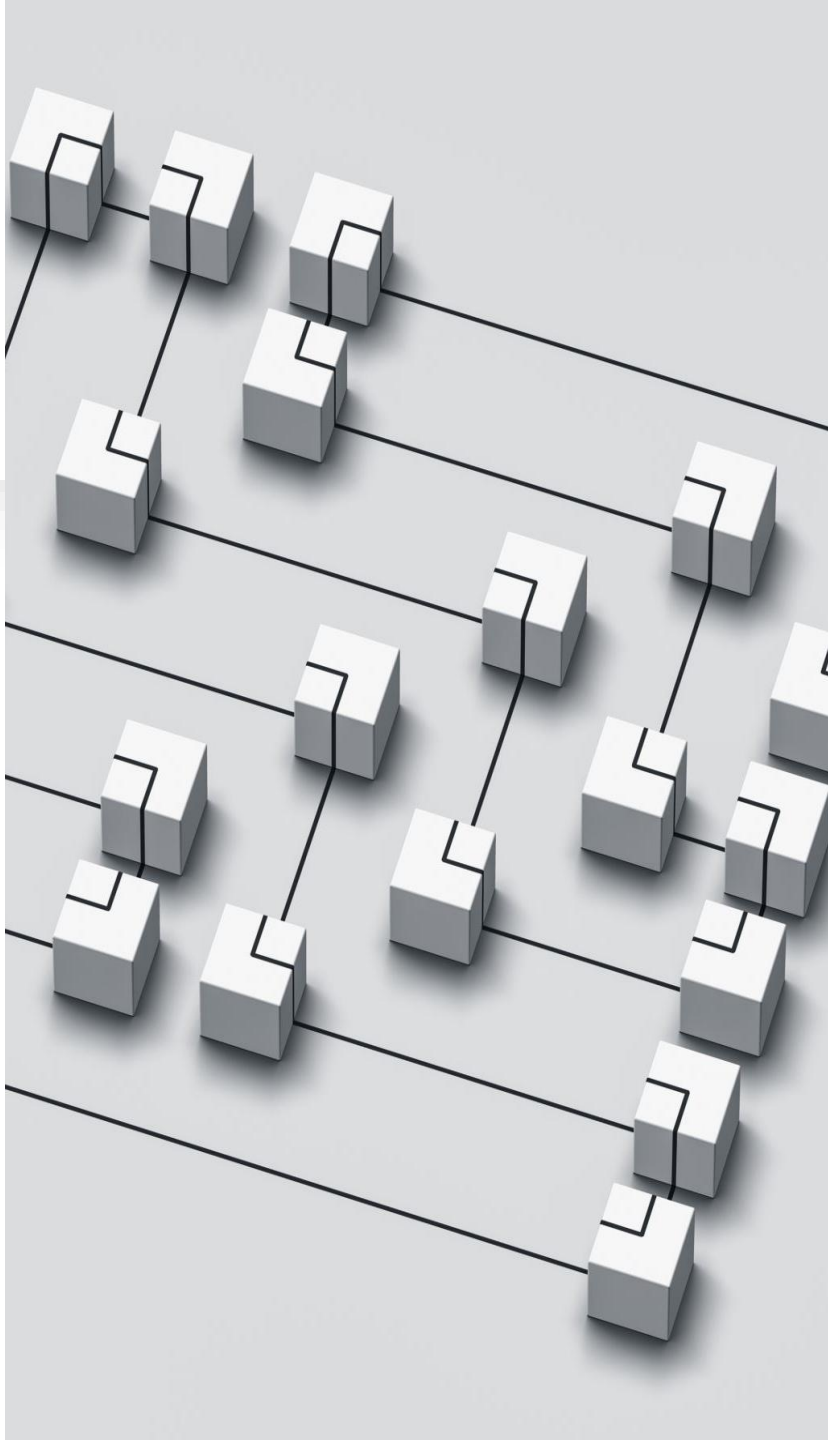
Git allows users to return to previous project states.

Each saved change is called a commit.

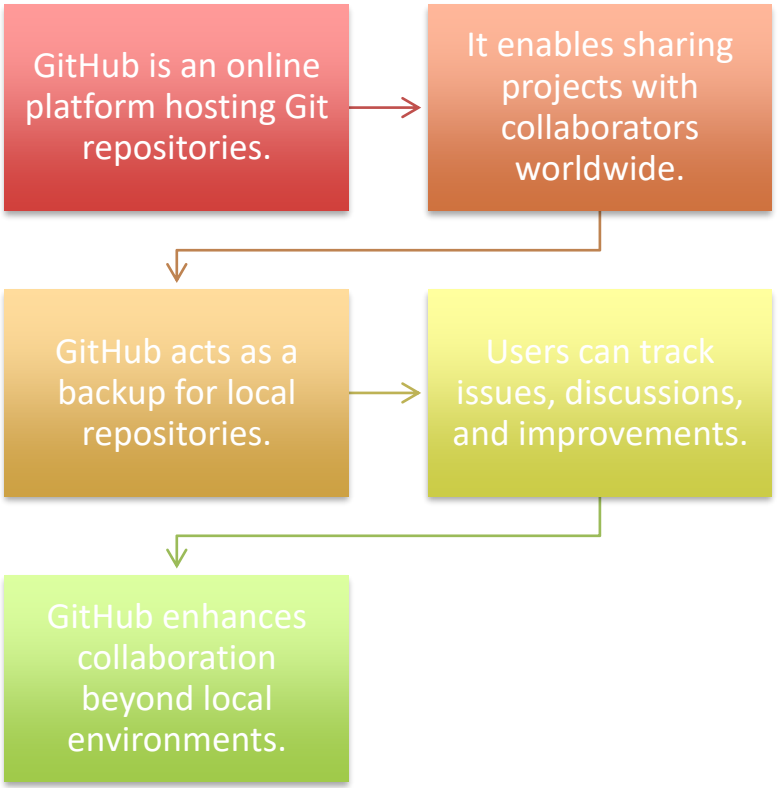
Git operates locally on your computer.

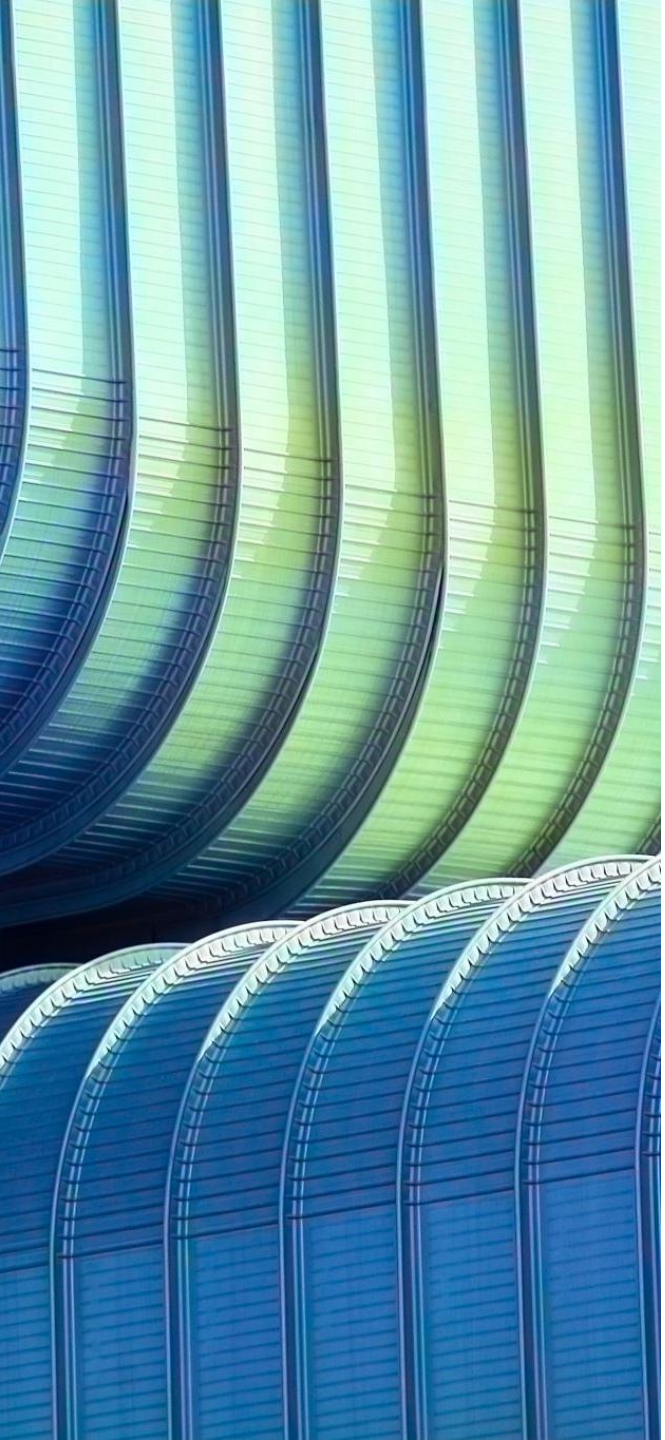
How Git Works Conceptually

- A repository contains all project files and history.
- Commits represent snapshots of the project at specific times.
- Each commit includes author, date, and description.
- Git tracks changes line-by-line across files.
- This system ensures complete project traceability.



What is GitHub?





Git vs GitHub

GIT IS THE VERSION CONTROL SOFTWARE ITSELF.

GITHUB IS A WEB-BASED HOSTING SERVICE.

GIT MANAGES CHANGES LOCALLY ON YOUR MACHINE.

GITHUB SYNCHRONIZES WORK BETWEEN MULTIPLE COLLABORATORS.

BOTH TOOLS WORK TOGETHER IN RESEARCH WORKFLOWS.

Why Use GitHub for Research

- GitHub preserves full history of analytical decisions.
- All changes are documented automatically over time.
- Collaborators can review, discuss, and suggest changes.
- Public repositories enhance transparency and openness.
- Private repositories support controlled collaborative research.

Using Git with R

RStudio integrates Git without requiring command-line usage.

Users manage Git actions through simple interface buttons.

Projects automatically track scripts, data, and outputs.

Git reduces risk of accidental code loss.

This approach suits beginners and experienced researchers.

Required Software Setup

R provides the programming environment for data analysis.

RStudio offers an integrated development interface.

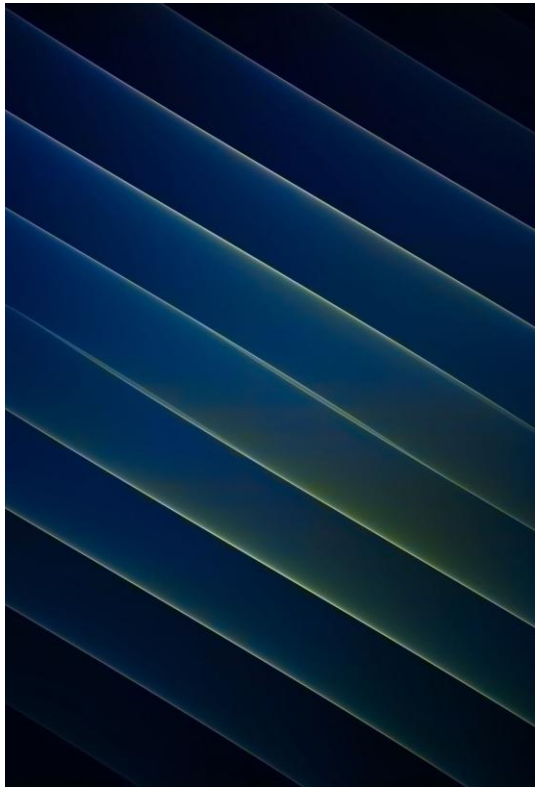
Git must be installed to enable version control.

A GitHub account enables online collaboration.

All tools are free and widely available.

Basic Git Workflow in R

- Create a new R project linked to Git.
- Modify scripts or data files as needed.
- Stage and commit changes with descriptive messages.
- Push commits to GitHub repository.
- This workflow preserves research history.



Collaboration Using GitHub

- Collaborators clone the shared repository locally.
- Each contributor commits their changes independently.
- Pull updates before making new modifications.
- Conflicts occur when same lines are edited.
- Git provides tools to resolve conflicts transparently.

Good Reproducible Practices

- Use clear folder structures for data and scripts.
- Write readable and well-commented analysis code.
- Use meaningful commit messages describing changes.
- Include README files explaining project structure.
- These practices improve long-term reproducibility.

Leveraging AI for Transparency and Reproducibility

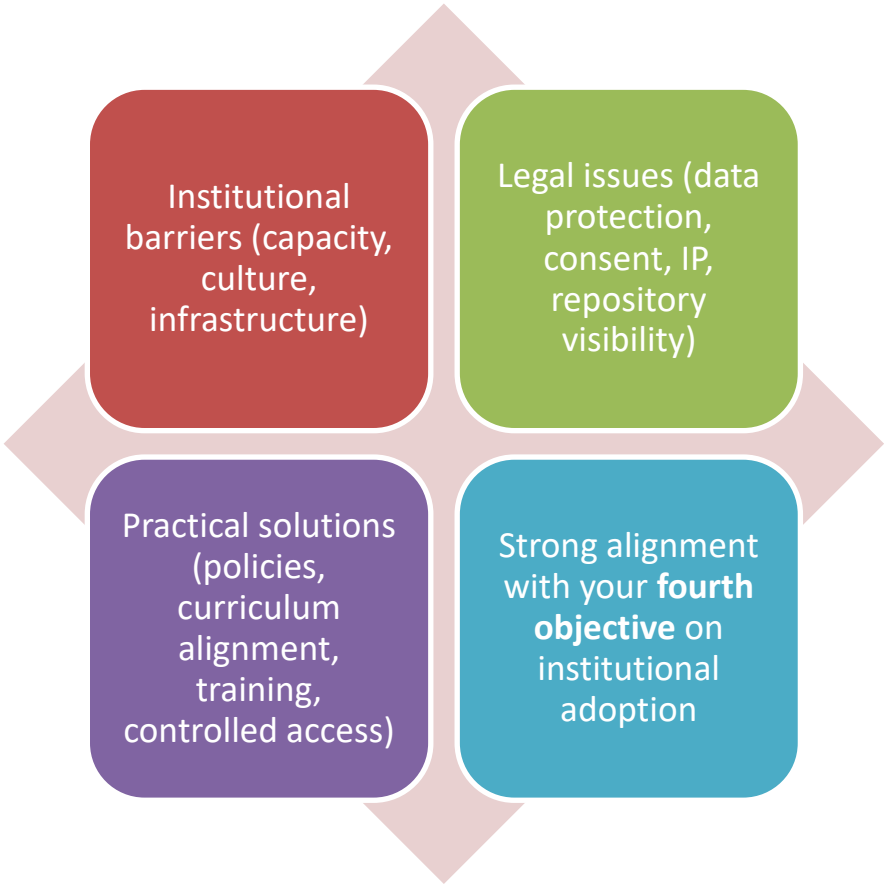
Practical, **non-technical** uses of AI in research (R code support, documentation, explanations)

Clear link to **transparency** (documentation, clarity, traceability)

Clear link to **reproducibility** (standardized workflows, readable scripts)

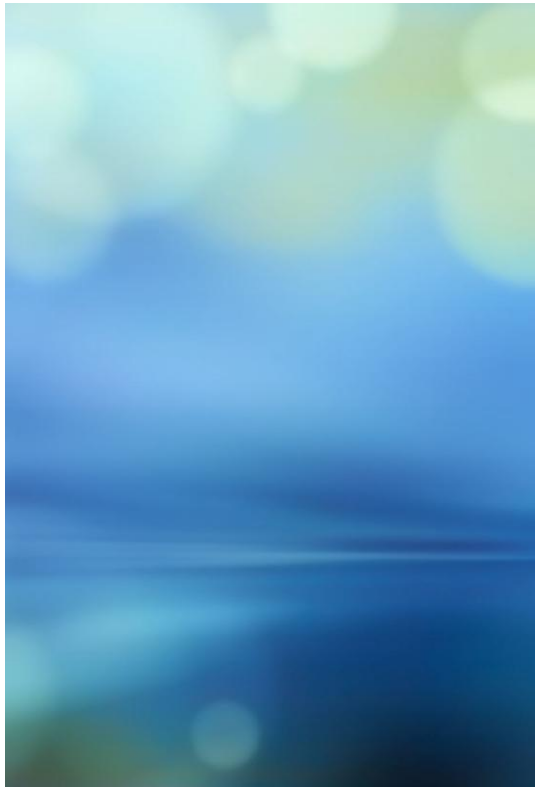
Explicit **ethical guidance**: disclosure, verification, data privacy, human responsibility

Institutional, Legal, and Ethical Contexts – Challenges and Solutions



What You Will Practice

- Create an R project with Git enabled.
- Connect the project to a GitHub repository.
- Make changes and commit them properly.
- Push updates to GitHub successfully.
- Simulate collaboration workflows safely.



Key Takeaways

- Transparency enables understanding and evaluation of research.
- Reproducibility ensures results can be independently verified.
- Git tracks analytical changes systematically.
- GitHub enables collaborative and transparent research workflows.
- R integrates seamlessly with Git-based practices.

Acknowledgements

This project was funded by the Berkeley Initiative for Transparency in the Social Sciences (BITSS), UNDER THE Center for Effective Global Action (CEGA)

Thank You

Compiled By:
Louis Damian
luisdamia6@gmail.com