

Unlocking Research Potential: Open Science and AI Tools for the Next Generation Scholar

Roots for Resilience Program (R4R)

SAKA Harvis Bamidele

PhD Student



November 21, 2025

□ TODAY'S ROAD MAP

□ What the program is about?

□ Topics covered

□ Main Takeaway & Research Relevance

□ open science

□ Application of LLMs: useful AI tools and how it helps
your research

□ Data management made easy

□ Conclusion



❑ WHAT THE PROGRAM IS ABOUT?

- Foundational Open Science Skills (FOSS) section : Tuesdays online / 11 – 1pm- Trainers: *Michele Cosi and Jeffrey Gillan*

FOSS: *Supports tools like the Open Science Framework (OSF) to help researchers manage and share their work transparently.*

□ ROOT FOR RESILIENCE (R4R)

- Root for Resilience (Practical section and some more.):
Thursdays in person /11am – 1pm- Data scientists and other campus partners;
 - Tina Johnson: The R4R Program Manager
 - The overall management of the program
- ***R4R:** Provides training and support to selected graduate students on open, reproducible science, computational infrastructure and AI tools to enhance research focused on environmental and societal resilience.*



COHORT MEET UP: ONCE A WEEK

Application:

<https://datascience.arizona.edu/r4r>

❑ TOPICS COVERED

- Open Science : Sharing research openly to accelerate discovery
- Data Management & Documentation: Organize your data for clarity
- How to Talk to Computers: effective digital communication.
- How to Talk to LLMs: Craft prompts that unlock powerful AI-driven support
- Reproducibility I: Version Control: Track changes and collaborate with Git.
- Reproducibility II Software Environments: controlled computing setups.
- Reproducibility III: Containers: Package your code and dependencies
- Remote Computing: HPC: Harness : large-scale scientific computing.
- Remote Computing Cyverse: Use cloud-based tools to analyze data.
- AI Models: Models to accelerate your research workflows.



OPEN SCIENCE

The movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of society, amateur or professional.

“Bear in mind that the wonderful things you learn in your schools are the work of many generations. All this is put in your hands as your inheritance in order that you may receive it, honor it, add to it, and one day faithfully hand it on to your children”. Albert Einstein.

❑ Open Science: Back in the Centuries

- ❑ In the 16th–17th centuries, science was often *closed*: discoveries were kept secret to protect prestige, patronage, or political advantage
- ❑ Scientists initially encoded findings in anagrams to avoid rivals stealing credit.
- ❑ Slows down progress!!!
- ❑ There have been Progressive shift: Digital Era, some Policies, National academies and societies promoted peer review, Public Libraries, Funders.



❑ CONFERENCES SUPPORTS OPEN SCIENCE

- Conferences embody *open science* by fostering *transparency* and *reproducibility*, sharing not just results, but data, methods, and workflows for *collection*
- *AIChE, Boston, 2022*





❑ SIX PILLARS OF OPEN SCIENCE

- ❑ *Open Access Publication:*
\$12,290, \$2,290
- **Subscription model** - the author pays a smaller fee (or no fee) for the article to be published.
- **Open Access model** - The author pays a larger fee to make the article freely available
- <https://lib.arizona.edu/research/share/open-access/support>

❑ OTHER PILLARS

- ❑ *Open Data: RE-data*: <https://data.library.arizona.edu/supported-platforms/redata/depositing-datasets>
- ❑ Open Educational source/Workshop: <https://carpentries.org/>
- ❑ Open Methodology: <https://github.com/> , <https://www.cos.io/initiatives/prereg>
- ❑ Open peer review: <https://prereview.org/>, <https://pubpeer.com/>
- ❑ Open source Software : <https://user.cyverse.org/> Cyverse UA

❖ *As open as possible and As closed as necessary!!!*

❑ LLMs: useful AI tools and how it helps research

❑ LLMs (Large Language Models): are *advanced AI systems trained to understand and generate human-like text (from internet) using massive datasets and billions of parameters.*

❑ **How they work?** LLMs don't understand like humans do. They predict what comes next in a sentence using math and probabilities. They don't have thoughts or feelings. *They write nonsense if not guided well.*

❑ **How can you use them?** You can use them for answering questions, writing essays, coding help, and more. *But you must be cautious because they can generate biased or false information if not used responsibly.* Be careful of *Hallucination!!!*.



❑ **Prompt Engineering: Role Playing**-"I want you to act as ..., "Start with simple, direct prompts before advancing to complex ones

❑ ROLE PLAYING

ChatGPT ▾

✦ Get Plus ✕

Act like a plumber: Tell me how to repair my broke pipe in the bathroom

Alright — plumber hat on 🧑🔧

Here's what you should do to repair a broken pipe in your bathroom:

1. **Shut off the water supply** — Find the nearest shutoff valve (usually under the sink or near the main line) and turn it off to stop the flow.
2. **Drain the line** — Open the faucets to release any remaining water.
3. **Locate the break** — Check for cracks, leaks, or bursts in the pipe.
4. **Cut out the damaged section** — Use a pipe cutter or hacksaw to remove the broken portion.
5. **Clean and dry the ends** — Make sure the remaining pipe edges are smooth and dry.
6. **Install a repair coupling or slip fitting** — For PVC, use new connectors with primer and cement; for copper, use a compression or push-fit (SharkBite) fitting.
7. **Tighten and secure** — Ensure a snug fit so there's no wiggle room.
8. **Turn water back on slowly** — Check for leaks while water pressure builds up.
9. **Seal and insulate** (if necessary) — Add insulation to prevent future damage, especially if it's a cold area.



An AI-generated country song is a smash hit, what does it mean for the industry?

November 13, 2025 at 2:37 PM MST

By Evan Hummel (Producer)



Summary

Full story

Sources

Summary

❑ Start with Simple statements and ramp gradually

❑ PROBLEMS FACED BY RESEARCHERS

❑ Researchers struggle to keep up with the volume of new publications and data: Chatbot that answers questions by searching the web in real time and summarizing results with citations(Perplexity)



❑ Finding the details in research paper// Skimming abstracts and full texts can be slow and cognitively draining, especially for interdisciplinary topics (Notebooklm)



❑ How much attention does a paper have. Especially with some Methods// Overwhelming volume of papers made it hard to spot underexplored areas or emerging trends (ResearchRabbit)



❑ Let's see how they work

❑ <https://lmarena.ai/leaderboard>

❑ <https://www.perplexity.ai/> Give me citations and references that supports the statement “As temperature increases Cation exchange capacity of the biochar decreases”

❑ <https://notebooklm.google/>

❑ <https://app.researchrabbit.ai/>

❑ ChatGPT: UA will soon subscribe for this!

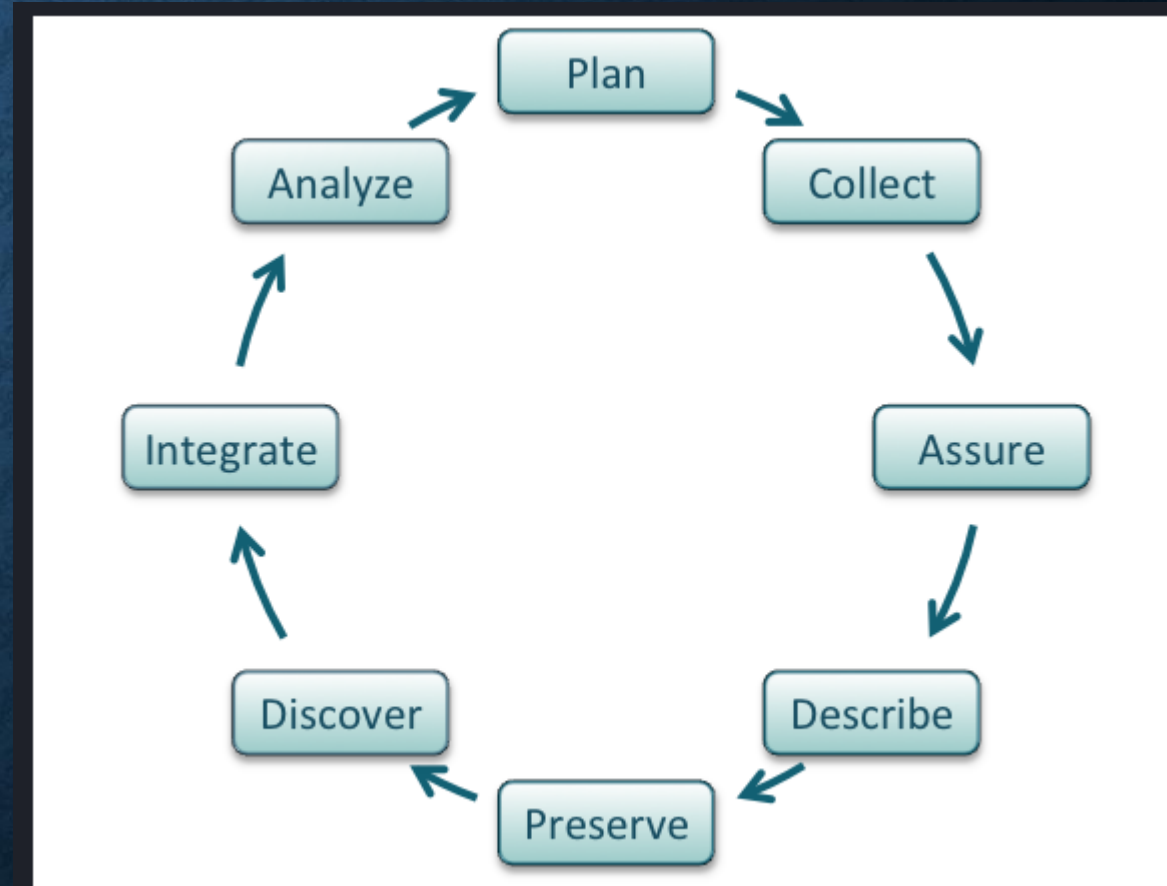
❑ <https://responsibleai.arizona.edu/students/student-guidelines-principles>

□ DATA MANAGEMENT PLAN (DMPs)

□ Data management is the set of practices that allow researchers to effectively and efficiently handle data throughout the data life cycle.

- Sponsors//funders usually ask for how data would be managed

□ Data life cycle



❑ WHAT CLASSIFIES AS DATA?

❑ Different types of data require different management practices. Here are some examples of what we can call Data.

❑ **Data Types:** **Text:** field or laboratory notes; survey responses; **Numeric:** tables, counts, measurements; **Audiovisual:** images, sound recordings, video; **Models, computer code;** **Instrument-specific:** equipment outputs

❑ Now that we've explored how data is classified, let's shift our focus to how it should be stewarded. The FAIR and CARE principles provide ethical and technical guidance for managing data responsibly.

□ FAIR AND CARE IN DMPs

- The **FAIR principles**: *Findable, Accessible, Interoperable, Reusable*-are foundational for designing a technically DMP.
- **In a DMP**, FAIR principles guide how data will be stored, documented, shared, and preserved to maximize its scientific utility
- The **CARE principles**: *Collective Benefit, Authority to Control, Responsibility, Ethics*-address the ethical dimensions of data governance, especially for Indigenous and marginalized communities
- **In a DMP**, CARE principles shape how consent, governance, and benefit-sharing are handled. They ensure that data management respects community rights and avoids extractive or exploitative practices

❑ For instance, if I consider Application of Biochar

- ❑ By aligning my DMP with both FAIR and CARE, I ensure that my biochar research, especially its application to *community soils*, is not only technically robust and *openly accessible*, but also *ethically grounded, culturally respectful*, and *responsive to the needs and rights of the communities involved*.
- ❑ FAIR helps you structure and share datasets via platforms like *ReDATA*.
- ❑ CARE guides how you *document consent*, benefit-sharing, and *community control over sensitive data*.

❑ PROBLEMS FACED BY RESEARCHERS

- ❑ Putting Data Management plans together for a research grant//funders
- ❑ Compliance with Funder and Journal Requirements
- ❑ Lack of Structure for Data Lifecycle
- ❑ Barriers to Reproducibility and Transparency
- ❑ Limited Visibility and Accessibility of Research Outputs

Let's navigate it

❑ <https://dmptool.org/plans>

❑ <https://ds-wizard.org/>

❑ CONCLUSION

- ❑ Open science fosters transparency, collaboration, and equitable access to knowledge, accelerating discovery and societal impact
- ❑ LLMs are changing tech.; The future: Faster (even faster than lawmakers), smarter, more ethical AI, deeply woven into daily life.
 - ❑ *“Artificial Intelligence(AI) must be used along with Human Intelligence(HI)”– Saka*
- ❑ DMP tools will be AI-powered, funder-aligned, and seamlessly embedded in open, collaborative research workflows

❑ APPRECIATION



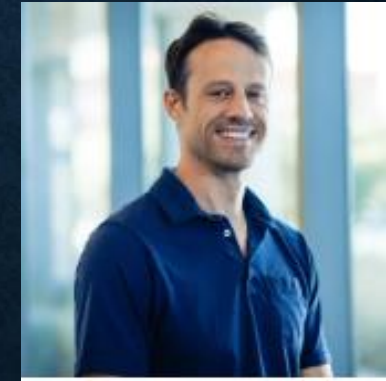
Prof. Kimberly Ogden



Cusi: Research Data Scientist



**Tina: R4R Program
Manager, Office of the
Chief AI Officer**



Jeffrey: Research Data Scientist

THANK YOU ALL FOR LISTENING!!!

