THE UNIVERSITY OF ARIZONA
Mel & Enid Zuckerman
College of Public Health

Integrating Open Science Principles and AI into Epidemiological Research

What is Roots for Resilience (R4R)?

Organizer

RESEARCH & PARTNERSHIPS
Data Science Institute

Arizona Institute
for Resilience

CYVERSE®

Roots for Resilience Program (R4R)

The Roots for Resilience program provides training and support to select graduate students on open, reproducible science, computational infrastructure and AI tools to enhance research focused on environmental and societal resilience.

R4R Goals

- Develop data science capabilities.
- Accelerate research projects .
- Build professional networks for addressing large-scale challenges and research questions.
- Develop new interdisciplinary collaborations across AIR, DSI, CyVerse, and other academic units.
- Develop a cohort among participants

Participation and Requirements

Who can participate?

- Any graduate student at the U of A with the support of their academic advisor – Complete R4R application

Requirements

- 2 hrs/week Foundation Open Science Skills workshop (online)
- 2 hrs/week In person cohort meetings
- 1 hr weekly cohort meetups
- Weekly journal on GitHub
- Capstone Project
- Departmental presentation

Content of the training

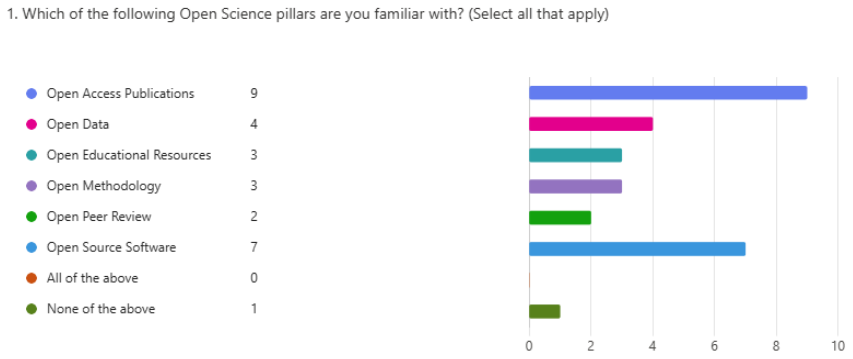
- **Introduction to Open science**
- **Data management & Documentation**
- How to talk to computers
- **How to talk to LLMs**
- **Reproducibility I: Version control**
- Reproducibility II: Software Environments
- Reproducibility III: Containers
- Remote computing: CyVerse
- Remote computing: High Performance Computing (HPC)
- AI models & where to find them (training your own AI models)

Week	Date	Content	Topic Overview	HackMD link
Week 1	Sept. 2	Intro to Open Science	- The big picture of what Open Science is and why do it. - How to apply the Open approach to various scientific processes and enhance your individual research.	Session 1 HackMD
Week 2	Sept. 9	Data management & Documentation	- Inferring on the importance of data and how your project can benefit from successful data management, discussing DMPs, storage, sharing and licences. - Emphasizing the role of effective documentation for research and projects in the age of Open Science.	Session 2 HackMD
Week 3	Sept. 16	How to Talk to Computers	- Learning of the command line, a more direct and powerful way to instruct your computer. - Introduction to the basics of the Unix Shell. - Why shell skills are useful for personal computing, cloud, and HPC.	Session 3 HackMD
Week 4	Sept. 23	How to Talk to LLMs	- Understanding how large language models (LLMs) work and how to effectively prompt them.	Session 4 HackMD

Survey results (n = 10)

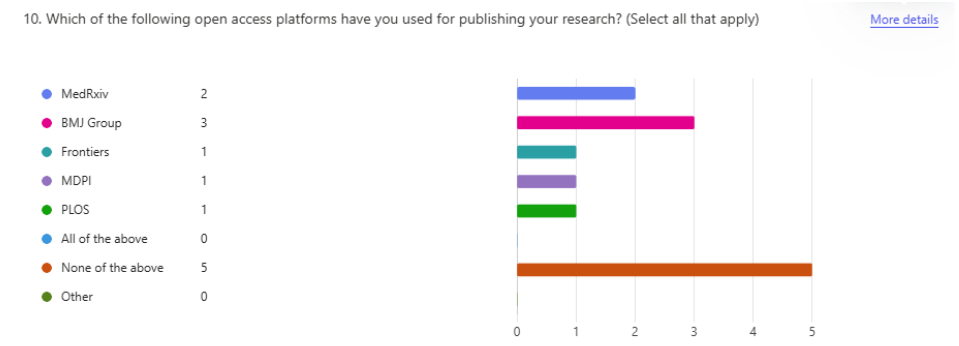
Familiarity/Usage:

- Overall open science knowledge/application.
- Open-source tools for data analysis and viz (R)
- Code sharing (GitHub)



Limited familiarity:

- Protocols and data sharing
- Collaborative writing
- Publications (preprints)
- UA Open Science Resources



Open science in epidemiological research

What is Open Science?

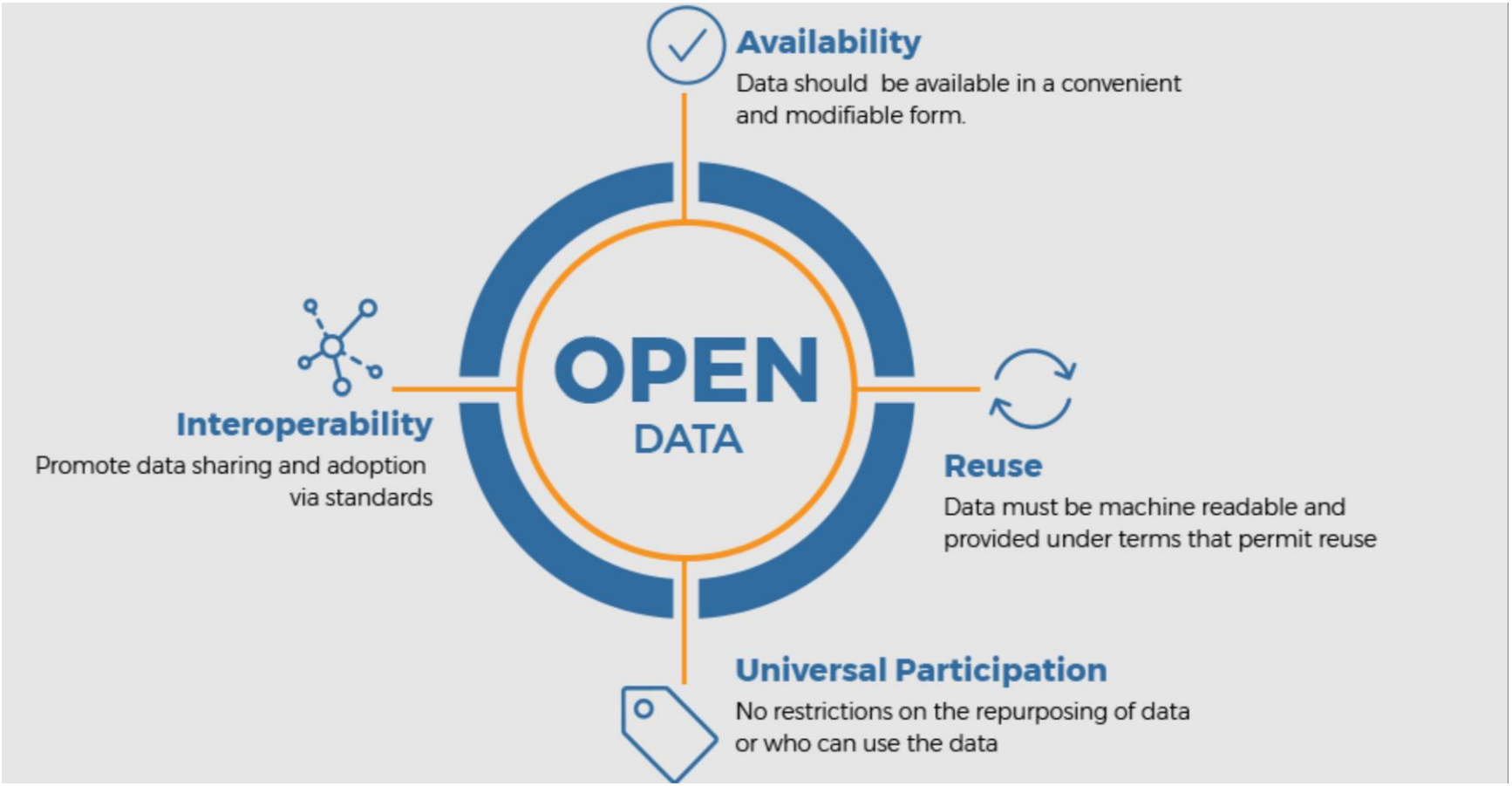
"Open Science is a collaborative and transparent approach to scientific research that emphasizes the accessibility, sharing, and reproducibility of data, methodologies, and findings to foster innovation and inclusivity." - ChatGPT

- FAIR data principles
  - Findable
  - Accessible
  - Interoperable
  - Reusable
- CARE data principles
  - Collective benefit
  - Authority to control
  - Responsibility
  - Ethics

Pillars of open Science:

OS Pillars: Open Access

OS Pillars: Open Data



OS Pillars: Open educational resources

- Aimed to benefit both the teachers and the students
- Educational materials that are free to use, customizable and easy to share.
- Example: Software carpentry

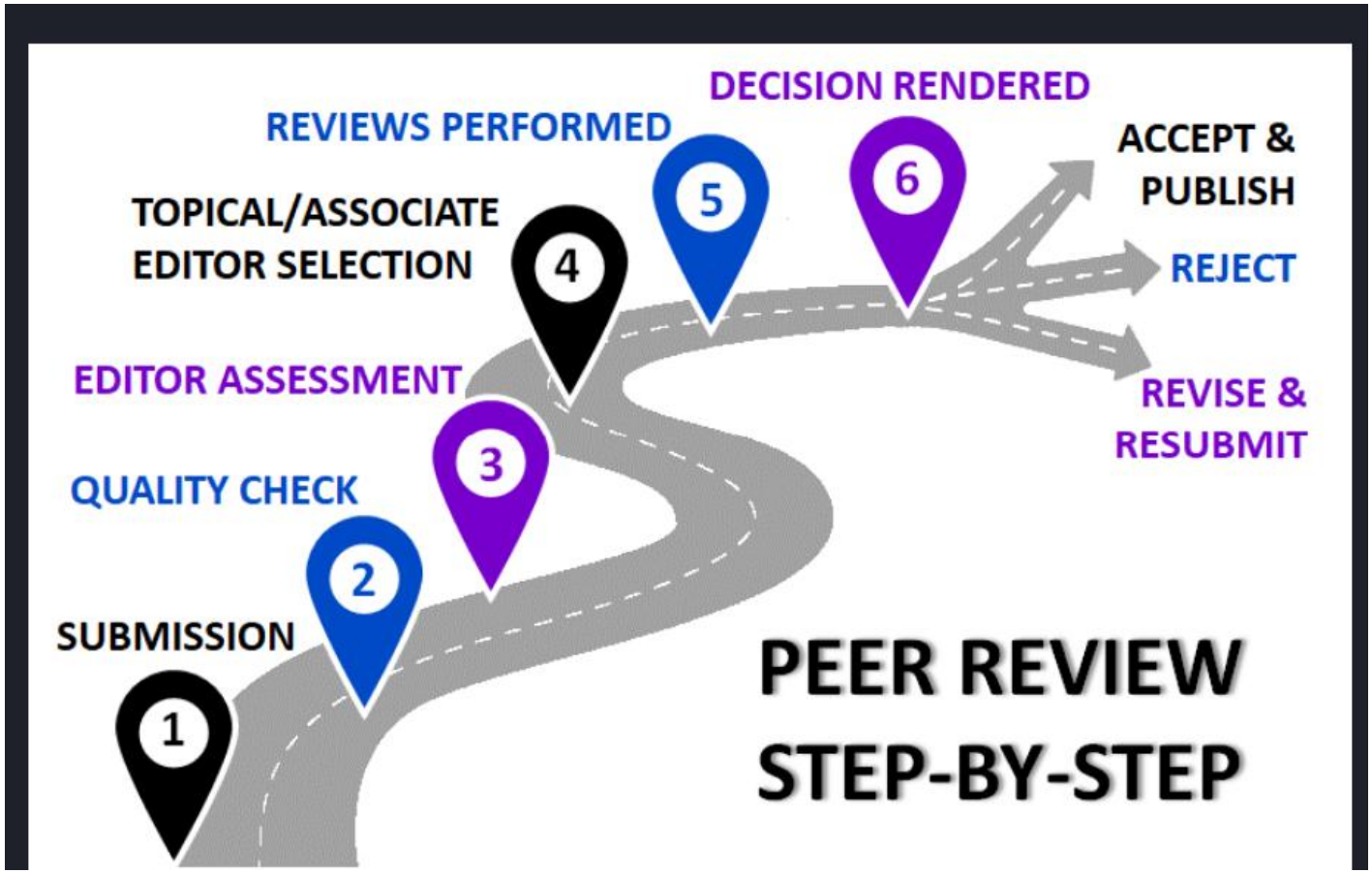
OS Pillars: Open Methodology

- Preregistration
- Sharing protocols and scientific data (Zenodo, OSF, PROSPERO, etc.)
- Data Management Plan
- Code sharing

DMP Tool

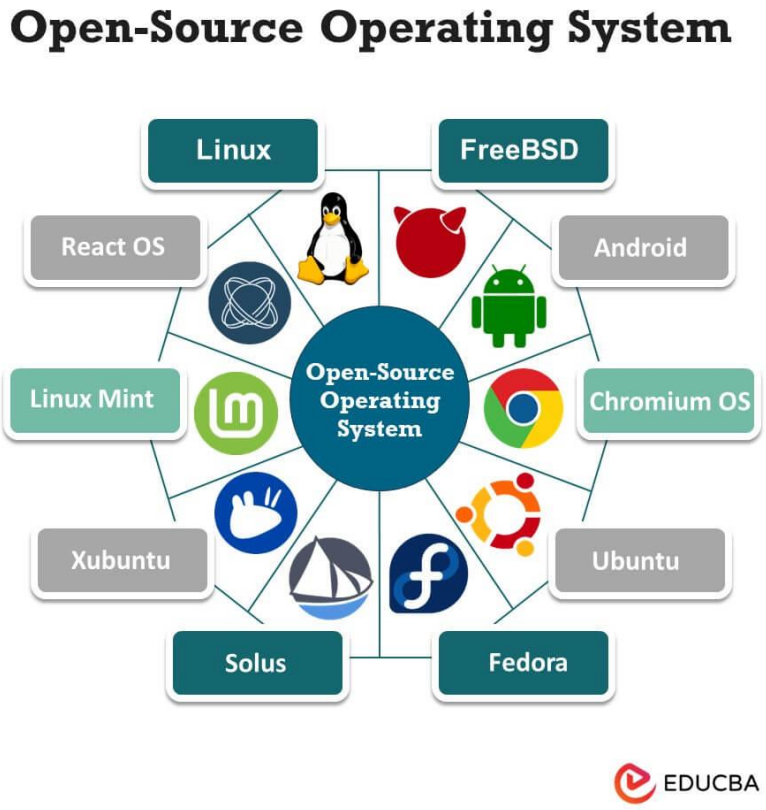
GitLab

OS Pillars: Open Peer Review



OS Pillars: Open-Source software

- Linux operating system and shell
- Python
- R
- Git/GitHub
- Conda
- Docker
- Cyverse



OS Pillars: Open-Source software

Reproducibility

- Version control using Git/GitHub
- Software environments: Conda/renv
- Docker containers

Source computing

- The University of Arizona Soteria.
- High Performance Computing (HPC)
- Cyverse Health

LLM chatbots for Open Science

Large Language Models (LLMs) have revolutionized how we interact with computers

ChatGPT

Gemini

Claude

Large Language Models

BCV

Application Layer

Copywriting
Jasper
copy.ai
Headlime
co:here
HyperWrite
Writesonic
Contenda
unbounce
copysmith

Coding
tabnine
MUTABLE AI
Codiga
co:here
GitHub Copilot
CODEGEN

Dev Tools
algolia
warp
Mintlify
cogram
Debuild
replit

Chat / Comms
MessageBird
Replier.ai
Sapling
FABLE

BizOps
viable
Enterpret
tabulate
Anecdote
OTHERSIDE AI
casetext
Dover

Infrastructure Layer

Model Creation
AI21labs
OpenAI
NVIDIA
Adept
EleutherAI
ANTHROP\C
Google AI

Hardware
habana
NVIDIA
SambaNova SYSTEMS
cerebras
GRAPHCORE

Fine Tuning
Google AI
aws
OpenAI
Hugging Face

Inference
OpenAI
Hugging Face

Communicating with LLMs

Prompt engineering.

Priming

CRAFT framework:

Context

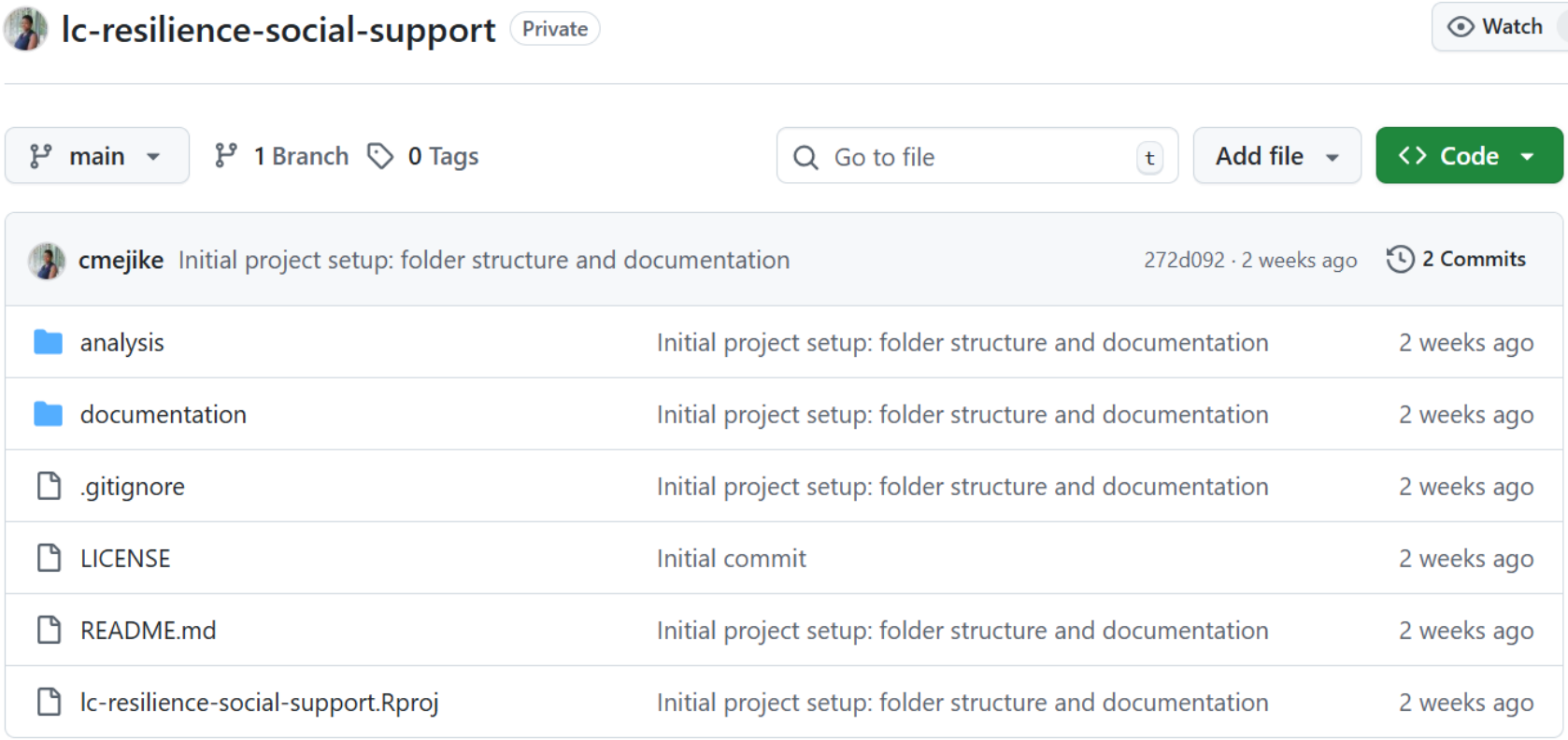
Role

Action

Format

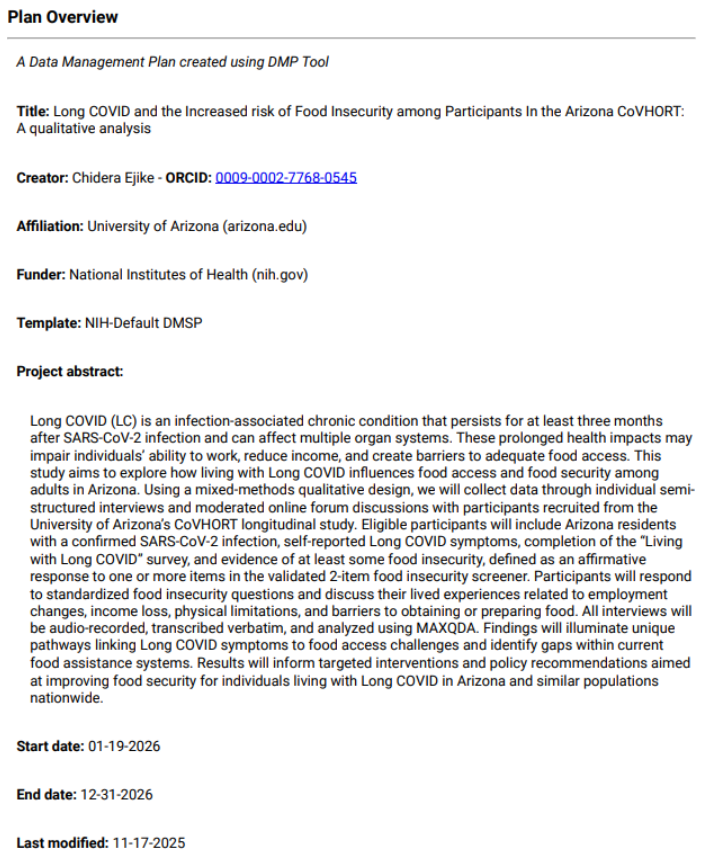
Tone

Integrating AI & Open Science
Asking AI the right questions: Step-by-step processes



Data Management Plan

I am writing a grant proposal to the National Science Foundation. Could you provide me a basic template for a data management plan (DMP) and please provide url links to resources that can help me with NSF DMP requirements.



Plan Overview

A Data Management Plan created using DMP Tool

Title: Long COVID and the Increased risk of Food Insecurity among Participants In the Arizona CoVHORT: A qualitative analysis

Creator: Chidera Ejike - **ORCID:** 0009-0002-7768-0545

Affiliation: University of Arizona (arizona.edu)

Funder: National Institutes of Health (nih.gov)

Template: NIH-Default DMSP

Project abstract:

Long COVID (LC) is an infection-associated chronic condition that persists for at least three months after SARS-CoV-2 infection and can affect multiple organ systems. These prolonged health impacts may impair individuals' ability to work, reduce income, and create barriers to adequate food access. This study aims to explore how living with Long COVID influences food access and food security among adults in Arizona. Using a mixed-methods qualitative design, we will collect data through individual semi-structured interviews and moderated online forum discussions with participants recruited from the University of Arizona's CoVHORT longitudinal study. Eligible participants will include Arizona residents with a confirmed SARS-CoV-2 infection, self-reported Long COVID symptoms, completion of the "Living with Long COVID" survey, and evidence of at least some food insecurity, defined as an affirmative response to one or more items in the validated 2-item food insecurity screener. Participants will respond to standardized food insecurity questions and discuss their lived experiences related to employment changes, income loss, physical limitations, and barriers to obtaining or preparing food. All interviews will be audio-recorded, transcribed verbatim, and analyzed using MAXQDA. Findings will illuminate unique pathways linking Long COVID symptoms to food access challenges and identify gaps within current food assistance systems. Results will inform targeted interventions and policy recommendations aimed at improving food security for individuals living with Long COVID in Arizona and similar populations nationwide.

Start date: 01-19-2026

End date: 12-31-2026

Last modified: 11-17-2025

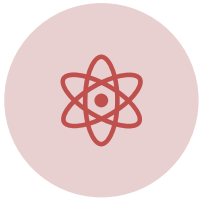
Challenges of Open Science in epidemiological research

- No formal “open science” training provided.
- Time-consuming
- Disinterest in learning new tools.
- Research system (quantity over quality)
- Preprint challenges
- Financial barriers
- Sensitive data barriers

Benefits of Open Science

RESEARCH REPRODUCIBILITY

LIMITS P-HACKING



SHARPENS SCIENTIFIC REASONING.

ENCOURAGES STREAMLINED WORKFLOWS

INCREASED EFFICIENCY

FACILITATES SECONDARY ANALYSES

EDUCATIONAL TOOL

Resources

University of Arizona Open Science resources for:

Planning research

Sharing research

Working with data

…and more

THE UNIVERSITY OF ARIZONA
Mel & Enid Zuckerman
College of Public Health

Psst... your data called... and it wants a plan...

*Help improve **data management and sharing practices** across the Epidemiology & Biostatistics Department.*

*Your responses will **inform tools, trainings, and presentations designed for you.***

take our 3-minute survey

University Libraries

Data Management & Sharing Plan Ambassador Program

Questions? Email Paulina Colombo,
Data Management & Sharing Ambassador

paulinacolombo@arizona.edu

THANK YOU!

PANEL DISCUSSION