Reproducing and Replicating Spatial Data Science

Joseph Holler, Peter Kedron, and Sarah Bardin

HEGSRR.github.io

Funding Support from NSF BCS-2049837, NSF OAC-1743184

Middlebury UCSB Arizona State University

Reproduction and Replication

Framing expectations

Workshop Agenda

Time	Topic	Presenters
15 min	What is R&R and why should you care?	Kedron, Bardin
25 min	Which open science practices are you familair with?	Holler, Bardin
15 min	What else do you want to know 1? (Discussion)	All
5 min	Break	All
10 min	What does the open science ecosystem look like?	All
30 min	What can open geographical science look like in practice?	Holler
5 min	How do you get started? Where can you contribute?	Kedron, Holler
15 min	What else do you want to know 2? (Discussion)	All

Central Objective:

1) Situate R&R in the geography and spatial data science

&

2) introduce selected research practices and resources you can use to make your work more reproducible

Tell us what you think

Please take 2mins to complete a short survey about open science research practices

Protocol

- Sort 11 open and reproducible research practices (ORRP)
  - Already using
  - Aware and interested in using
  - Unaware or uninterested
- 10 likert technology adoption questions
- Share deidentified data
- Follow-up surveys (immediate, one year)

Defining Reproducibility and Replicability

A brief definition for spatial data science

Reproduction and Replication

(Schmidt 2009, Gomez et al. 2010, Barba 2017, Christensen et al. 2019, NASEM 2019)

TABLE 1. Types and Purpose of Replication

Reproduced from Christensen et al. (2019, p159, Table 9.1)

Compared to original study	**Focused on *repeating procedures***	**Focused on *introducing differences***
***Same* data**	Verification	Reanalysis
***Different* data**	Direct Replication	Extension

Reproduction and Replication

(Schmidt 2009, Gomez et al. 2010, Barba 2017, Christensen et al. 2019, NASEM 2019)

TABLE 1. Types and Purpose of Replication

Reproduced from Christensen et al. (2019, p159, Table 9.1)

Compared to original study	Focused on *repeating procedures*	Focused on *introducing differences*
***Same* data**	Verification	Reanalysis
***Different* data**	Direct Replication	Extension

Reproduction

Same data, same procedure, same results, same context

Reproduction and Replication

(Schmidt 2009, Gomez et al. 2010, Barba 2017, Christensen et al. 2019, NASEM 2019)

TABLE 1. Types and Purpose of Replication

Reproduced from Christensen et al. (2019, p159, Table 9.1)

Compared to original study	Focused on *repeating procedures*	Focused on *introducing differences*
***Same* data**	Verification	Reanalysis
***Different* data**	Direct Replication	Extension

Reproduction

Same data, same procedure, same results, same context

Replication

New data, similar procedure, similar results, same or new context

Veridical Spatial Data Science

Munafo et al. (2016), Kedron et al. (2020), Yu & Kumbier (2020)

Principled inquiry to extract reliable and reproducible information from spatialtemporal data, with an *enriched technical language* to communicate and evaluate empirical evidence in the context of human decisions, domain knowledge, and geographic confounds; *supported by a system of external validation and evidence accumulation based on the purposeful replication of findings across space and time*.

(Adapted from Kedron and Bardin 2021, Yu and Kumbier 2020)

Missing Replication Studies

Geographers are thinking about R&R, but not attempting replications

BCS-2049837

Reproducibility

81% Consider reproducibility of their own work

71% Discuss reproducibility with colleagues

53% Consider reproducibility during peer-review

Replicability

74% Consider replicability of their own work

65% Discuss replicability with colleagues

59% Consider replicability during peer-review

14%

Reported attempting a reproduction
(**7%** attempt publishing)

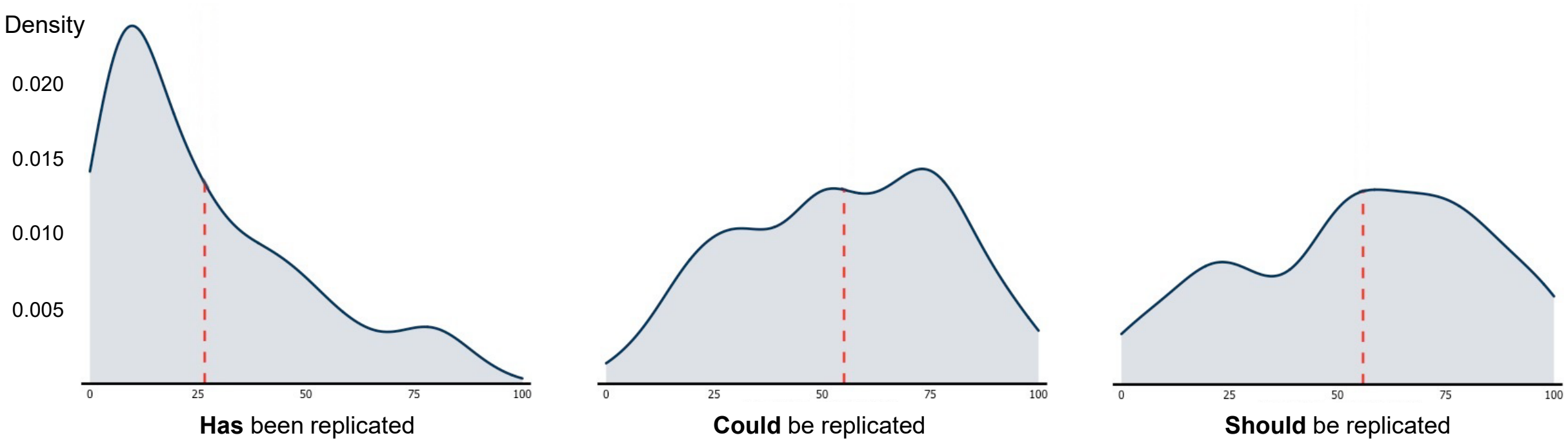
31%

Reported attempting a replication
(**21%** attempt publishing)

Missing Replication Studies

What percentage of geographic research do you believe …

BCS-2049837



2023:
The Year of Open Science

Sarah Bardin Riffs on R&R in the Policy Environment

What percentage of geographic research do you believe …

Sarah Bardin Riffs on R&R as a matter of Practice

What percentage of geographic research do you believe …

Our R&R Related Resources

Munafo et al. (2016), Kedron et al. (2020), Yu & Kumbier (2020)

hegsrr.github.io

hegsrr.github.io/Workshop-SDSS-2023/

- 5 Peer-reviewed Publications
- 8 Reproduction and Replication Studies
- 2 Surveys of Researcher Practices
- **Reproducible Project Repository Template**
- Manual In Development
- 2 Course Syllabi
- 9 RAs Mentored
- ~50 Students Engaged in R&R Studies

Survey Results

The reproducible research practices you identified as interesting and important

Link to Report

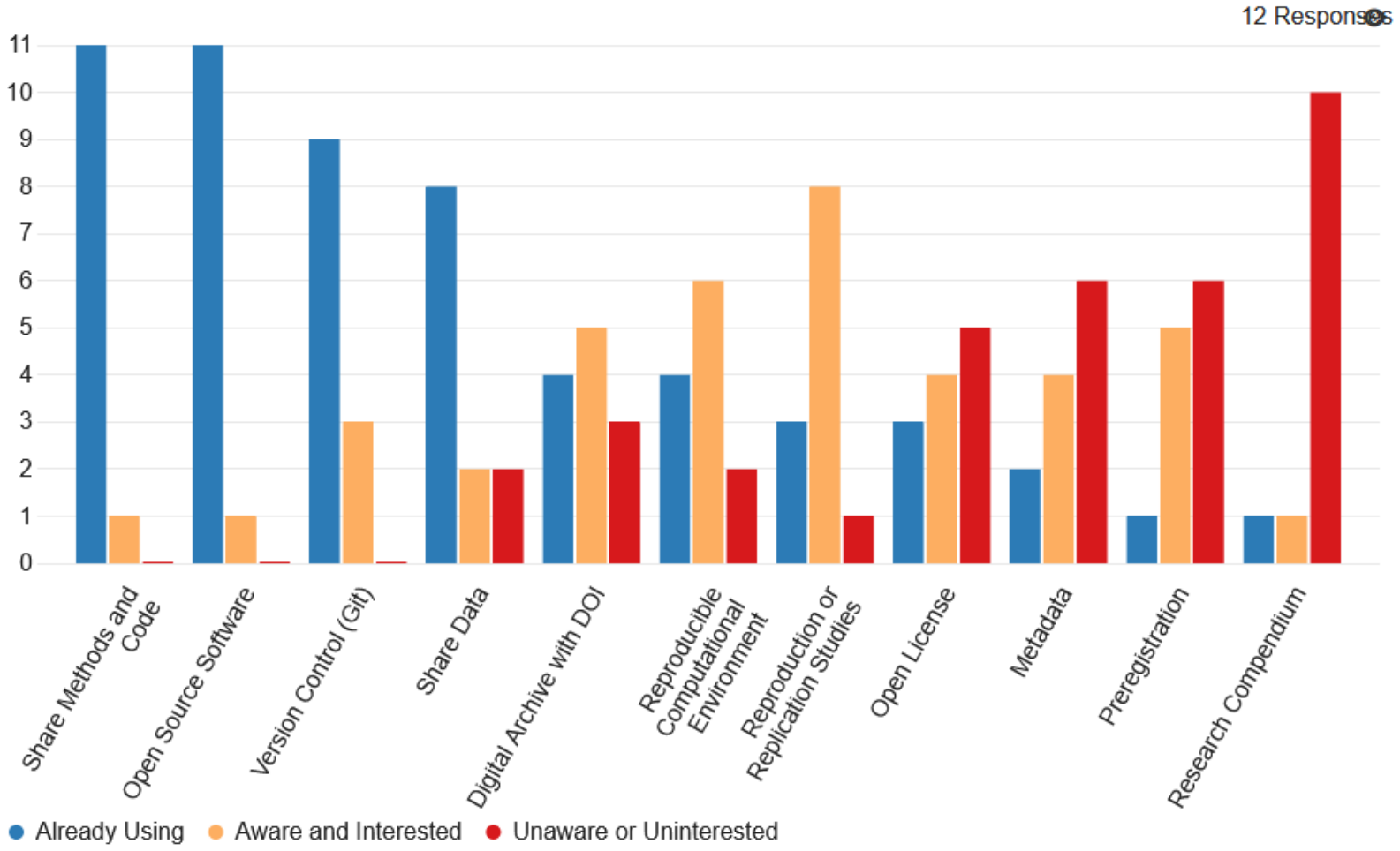
Open and Reproducible Research Practices

A quick review of practices in light of the survey results

Survey Results: You

Goal: incremental progress as individuals and scientific community

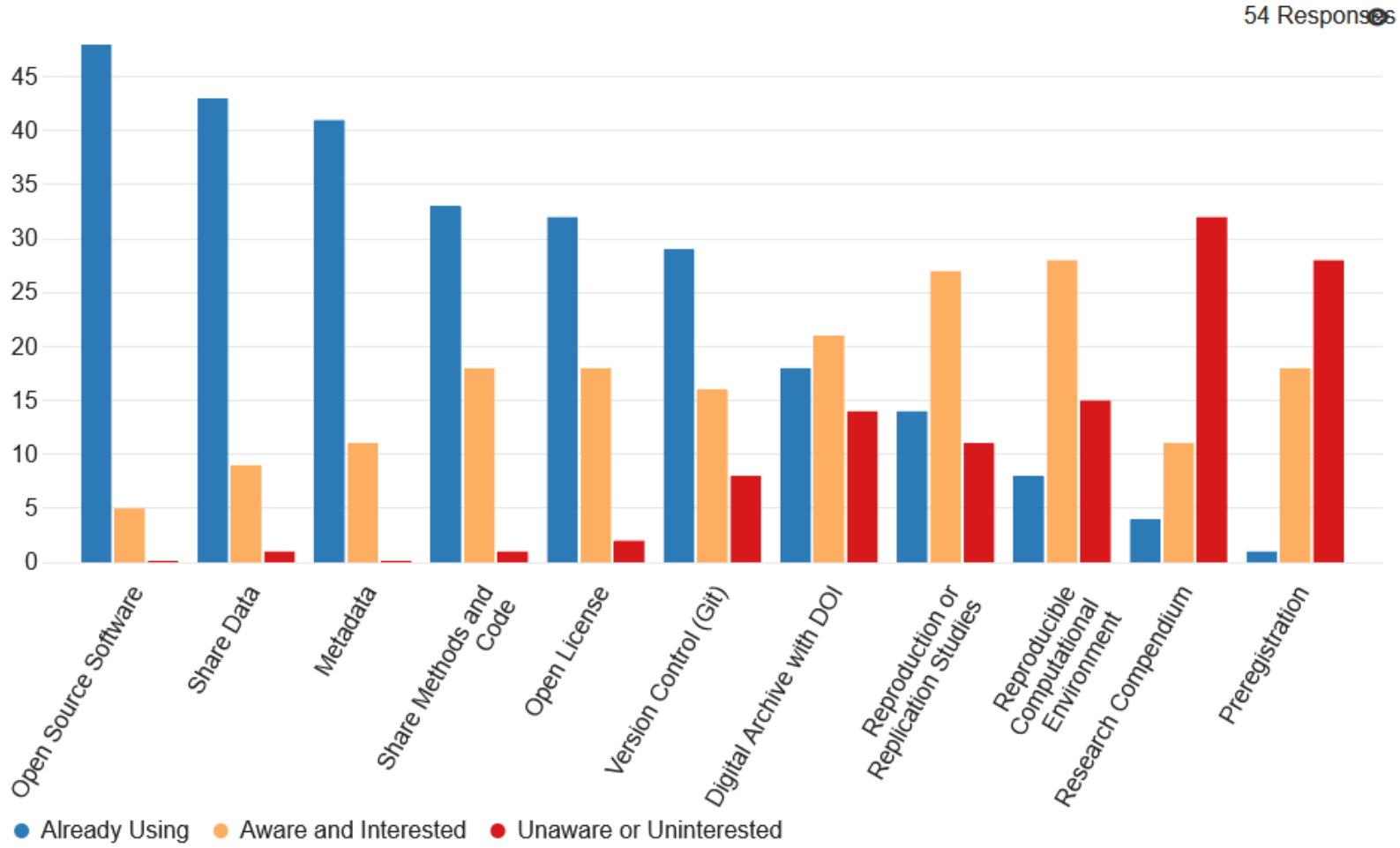
Be kind to your future self!



Results: Spatial Data Science Symposium

Goal: incremental progress as individuals and scientific community

Be kind to your future self!



Share Methods and Code

Rmarkdown, Jupyter Notebook, Scripts, Models, Protocols

Do you share a complete description of your methods?

Open Source Software

Python, R, PostGIS, GeoDA, QGIS

Do you use and cite research software with (re)distributable source code?

Version Control

Git, GitHub, GitLabs, OSF projects

Do you manage and track changes
to your study design, data, and code?

- Best at tracking one line of plain text
- GitHub integrates with Overleaf, OSF, Webpages

Share Data

Do you make the data for your study readily available in the most complete and unmodified form permissible by law and ethical protocols?

Digital Archive with DOI

DOI: Digital Object Identifier

Are all the components of your study digitally archived for long-term preservation, and labelled and linked with a persistent digital object identifier?

Reproducible Computational Environment

Docker container, public cyberinfrastructure, environment metadata

Do you provide access to your computational environment or sufficient information about your environment such that others can recreate it?

Reproduction and Replication Studies

Do you attempt and share
reproduction or replication studies?

Open Licensing

Creative commons, BSD-3, GPU, MIT…

Do you license your research products to allow others to use, modify, and redistribute them?

Metadata

Dublin Core, ISO 191**, Federal Geographic Data Committee (FGDC)

Do you provide information about your study and each of its components in a standardized format?

Preregistration

OSF, AsPredicted, Registered Reports

Do you publicly register your hypotheses and research design before conducting your work?

Research Compendium

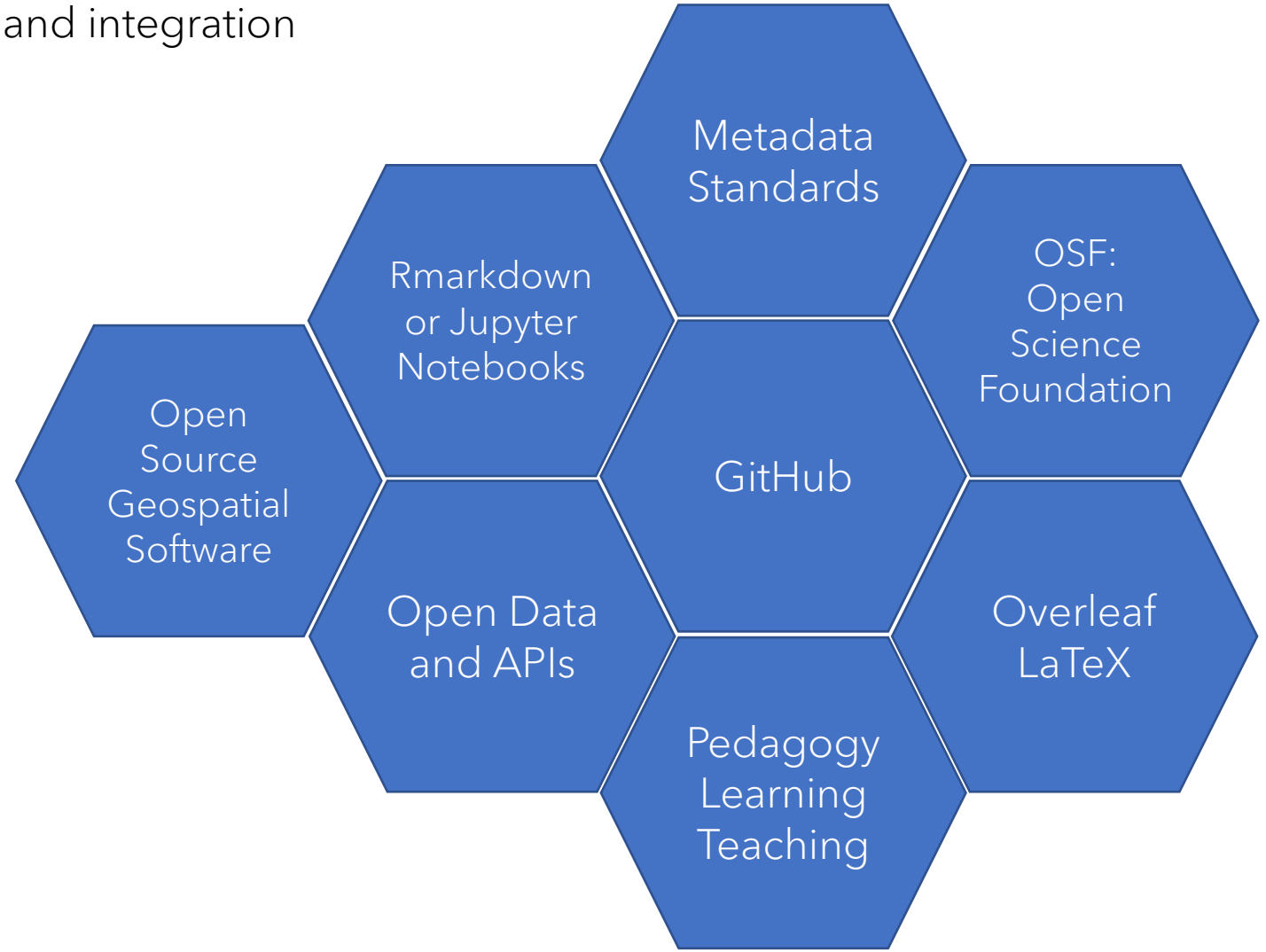
HEGSRR Template, TIER Protocol, WORCS, o2r

Do you collect *all* components of your study together in a directory organized with consistent structure and relative links?

Open Science Ecosystem

Multiple points of entry and integration

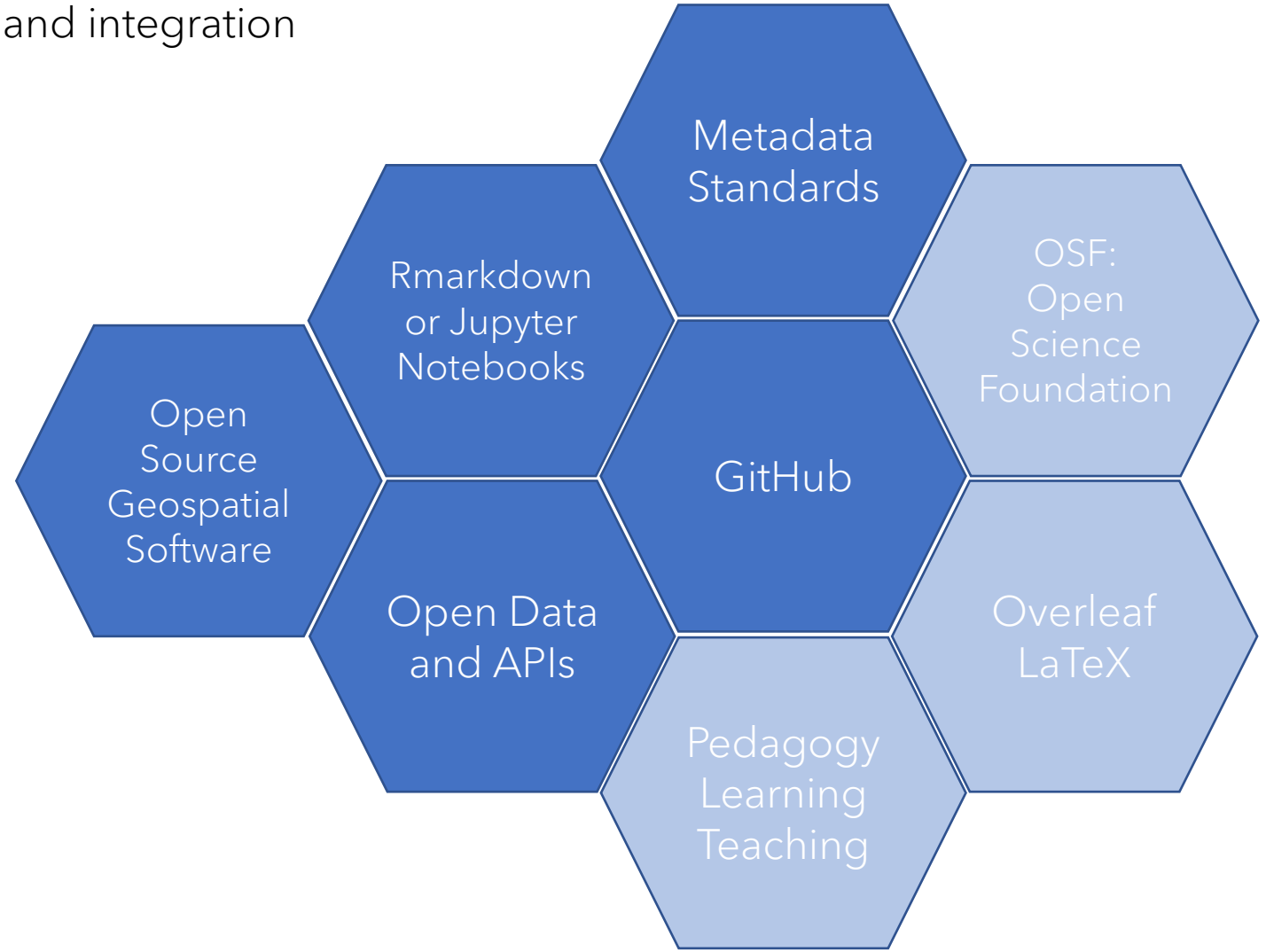
Where are you in your own research project and training?



Open Science Ecosystem

Multiple points of entry and integration

In one undergraduate course…



HEGS-RR Infrastructure

A project-based demonstration of our infrastructure for reproducibility and replicability

Questions about Infrastructure?

We invite questions, feedback, and discussion about R&R in spatial data science

Open Discussion

We invite questions, feedback, and discussion about R&R in spatial data science

Discussion prompts

and invitation to collaborate…

1. Can you share any *successes*, *advice*, or *best practices* introducing reproducibility and replicability in your own scholarship (research or teaching)?

2. What *barriers* do you perceive to adopting open and reproducible research practices in your own scholarship?

3. Could any *resources, changes,* or *incentives* help overcome those barriers?

Thank You

Our R&R Related Resources

hegsrr.github.io

/Workshop-UCSB-2023

- 5 Peer-reviewed Publications
- 8 Reproduction and Replication Studies
- 2 Surveys of Researcher Practices
- **Reproducible Project Repository Template**
- Manual In Development
- Course Syllabi
- 9 RAs Mentored
- ~50 Students Engaged in R&R Studies