2021 Workshop: Reproducibility in Geospace Science

Long title
Reproducibility in Geospace Science: Best practices for Data Stewardship
Conveners
InGeO team - Asti Bhatt, Ashton Reimer, Leslie Lamarche, Todd Valentic, Pablo Reyes
Tomoko Matsuo
Ryan McGranaghan
Description

Primary Objective: This workshop endeavors to advance discussions on computational reproducibility in CEDAR science, putting emphasis on the challenges posed by access to and usage of data and diverse stake-holders' needs.

Reproducing scientific results is a key component to the scientific method. For observational sciences, reproducing results is challenging, especially when observing largely nonlinear natural phenomena. However, in the era of widely available open data and software analysis tools, we need to also ensure computational reproducibility of research results. Towards that, it is critical to identify the needs of instrument developers, data providers, software developers, and scientific users in the CEDAR community to make research results computationally reproducible. This includes issues such as data collection and curation, data distribution and redistribution, data traceability and archive, credit attributions to data and software providers, incompatible licensing across different software and datasets, journal and funding agency requirements, and issues of user privacy and intellectual properties.

We invite robust discussions focused on the changing landscape of data and software publishing requirements for journals, licensing and citations of data and software, tracking users of scientific datasets and software for funding requirements, utility of data repositories and doi generation. We encourage the participation of students and early career researchers, especially in sharing challenges they have encountered attempting to perform reproducible research when using software

and/or datasets. The session will be organized in short 5-minute presentations and a panel discussion along with breakout rooms.

These discussions will inform the content of a whitepaper that will serve as an update to ‘Essential Best Practices for the Geospace Community Concerning Reproducible Research, Open Science, and Digital Scholarship’ (authored in 2018). This white paper will serve as a starting point for broader discussion specifically among ground-based observations within the CEDAR community on data and software stewardship towards creating reproducible results.

Agenda

Link to workshop whitepaper notes (pdf)

13:00-13:10: Introduction

Session Convenors

Introduction to the session and review the format and goals.

13:10-13:30: Data distribution and Licensing

Topic expert: Kathryn McWilliams, SuperDARN Data Distribution and Licensing

- Data availability requirements of funding agencies and journals. Data licensing. Data distribution and tracking. Long-term data distribution

13:30-13:50: Common data repositories

Topic expert: Bill Rideout, CEDAR Madrigal Database

- Existing data repositories for non-NASA data products, how are they used? Do they serve all the needs of the CEDAR community? . Metadata and file format standards. Can we enforce certain community standards with data management plans?

13:50-14:10: Data citation and attribution

Topic expert: Lan Jian, NASA SPDF

- Accessibility of data for users and how to comply with journal data policies. Recognition/credit for data providers. The need for a common community

“Rules of the Road”. Incentive structures for proper attribution and proper citation.

14:10-14:30: FAIR geospace data

Topic expert: Liam Kilcommons, AMGeO

- How are we doing with the different aspects of Findable, Accessible, Interoperable, Reusable (FAIR). Why do we need FAIR data? Are FAIR standards practical for CEDAR data?

14:30-14:50: Learning from other disciplines

Topic expert: Kenton McHenry, GEOCODES

- How other disciplines deal with data and what can we learn from that.

14:50-15:00: Conclusion

Session Convenors

- Wrap up, next steps, and closing remarks.

Justification

This workshop will address Strategic Thrust #6 "to manage, Mine and Manipulate Geoscience Data and Methods."

Reproducing computational results requires robust community infrastructure and adoption of best practices towards managing both data and software. The era of FAIR (Findable, Accessible, Interoperable, Reproducible) data is already upon us, yet we are not fully prepared to adhere to that practice. This is especially challenging in geospace science where we need distributed instrumentation and specialized computational software developed with funding secured through competitive processes. to carry out research. We hope to hear from instrument operators, data providers, and software developers in the geospace community on their challenges to balance various aspects of providing good data and software products and ensuring sustainability of data and software.

View PDF