Software Practices for Reproducible Science

Anshu Dubey and Gregory R. Watson

2024 ACM Conference on Reproducibility and Replicability (ACM-REP)

Presented by

COLABS: Collaboration for Better Software for Science

In collaboration with

With prior support from

Slides, late-breaking updates, etc. at:
https://bssw-tutorial.github.io/
and click the link for today's tutorial

License, Citation and Acknowledgements

License and Citation

- **The requested citation the overall tutorial is:** Anshu Dubey and Gregory R. Watson, Software Practices for Reproducible Science tutorial, in 2024 ACM Conference on Reproducibility and Replicability (ACM-REP), Rennes, France and online, 2024. DOI: 10.6084/m9.figshare.26019469**.**
- Individual modules may be cited as *Speaker, Module Title*, in *Tutorial Title*, …

Acknowledgements

- This work was supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR), and by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.
- This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Next-Generation Scientific Software Technologies (NGSST) program.
- This work was performed in part at the Argonne National Laboratory, which is managed by UChicago Argonne, LLC for the U.S. Department of Energy under Contract No. DE-AC02-06CH11357.
- This work was performed in part at the Lawrence Livermore National Laboratory, which is managed by Lawrence Livermore National Security, LLC for the U.S. Department of Energy under Contract No. DE-AC52-07NA27344.
- This work was performed in part at the Los Alamos National Laboratory, which is managed by Triad National Security, LLC for the U.S. Department of Energy under Contract No.89233218CNA000001
- This work was performed in part at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.
- This work was performed in part at Sandia National Laboratories. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

About Us

- Anshu Dubey, ANL
- Gregory R. Watson, ORNL

Anshu
she/her

Greg W
he/him

- Anshu
  - Lead PI for COLABS: Collaboration for Better Software for Science
  - Application Engagement Lead for the RAPIDS SciDAC Institute
  - Member of the IDEAS Productivity Project: http://ideas-productivity.org
- Greg
  - PI for CORSA: Center for Open-Source Research Software Advancement (https://corsa.center)
  - PI for the Neutrons Data Interpretation Platform project (ORNL)
  - Co-PI for the NZ-ARMADA project (ORNL)

Building an Online Community

https://bssw.io

- **New community-based resource for scientific software improvement**
- A central hub for sharing information on practices, techniques, experiences, and tools to improve developer productivity and software sustainability for computational science & engineering (CSE)

Goals

- Raise awareness of the importance of **good software practices** to scientific productivity and to the quality and reliability of computationally-based scientific results
- Raise awareness of the **increasing challenges** facing CSE software developers as high-end computing heads to extreme scales
- Help CSE researchers **increase effectiveness** as well as leverage and impact
- **Facilitate CSE collaboration via software** in order to advance scientific discoveries

Site users can…

- **Find information** on scientific software topics
- **Contribute new resources** based on your experiences
- Create content tailored to the unique needs and perspectives of a focused scientific domain

Follow BSSw

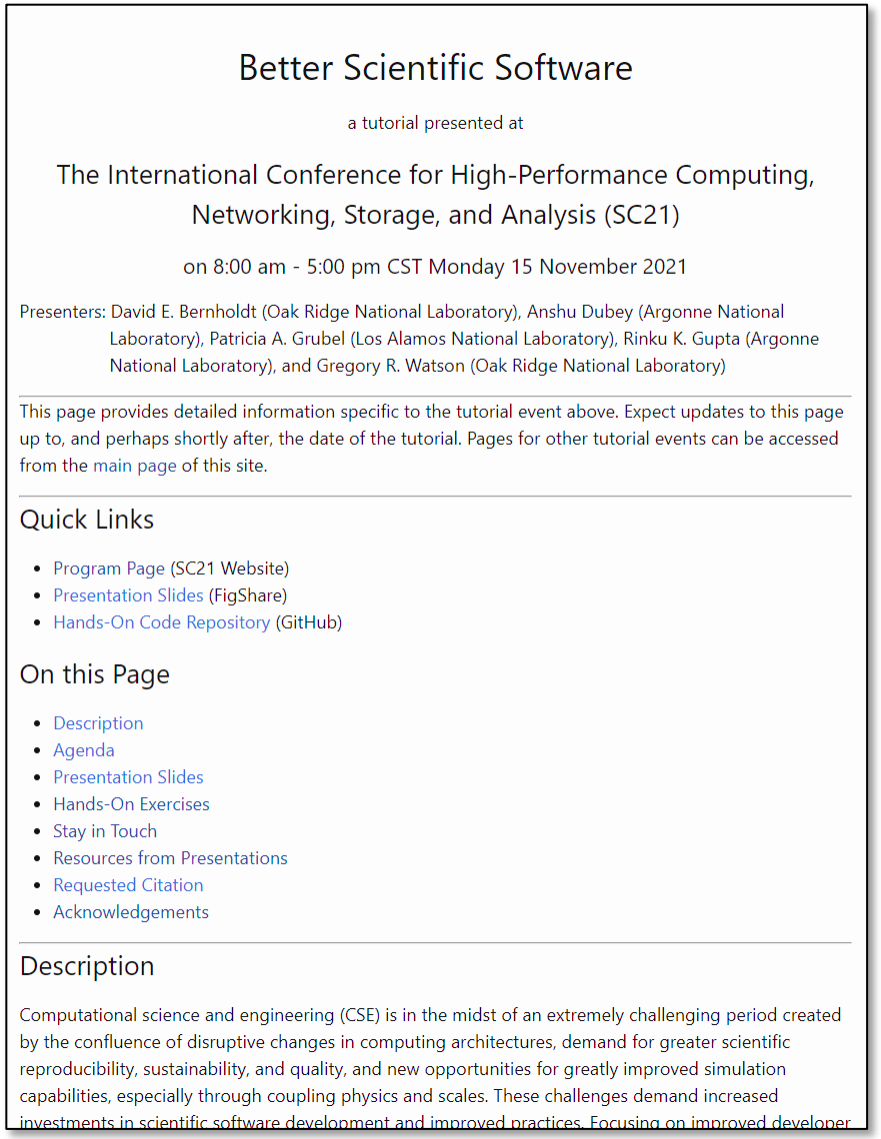
- BSSw Digest: https://bssw.io/pages/receive-our-email-digest
  - Updates on BSSw content
    - New blog posts, events, and resources
    - BSSw Fellowship
  - Typically 1-2 messages per month
  - Also: RSS feed: https://bssw.io/items.rss

The Importance of Naming

- Computing is rife with terminology that many consider harmful and exclusionary
  - Examples include: whitelist/blacklist, master/slave, and master (standalone)
- We support efforts to replace such language with more inclusive language
- In this tutorial, we strive to use inclusive language
  - Example: we use “main” for the default git branch, even where outside sources we reference may use “master”
- We welcome suggestions for further improvements in our tutorial
- Additional information:
  - The Inclusive Naming Initiative
  - The BSSw.io resource on inclusive naming provides some additional context and links

BSSw Tutorial Web Site

- https://bssw-tutorial.github.io/ is the one URL you need to find all the resources for this tutorial
- Each tutorial event has its own page
- Each tutorial page is considered archival
  - All of the materials used in that tutorial (or links to them)
  - Materials may be updated if we find mistakes



Better Scientific Software

a tutorial presented at

The International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC21)

on 8:00 am - 5:00 pm CST Monday 15 November 2021

Presenters: David E. Bernholdt (Oak Ridge National Laboratory), Anshu Dubey (Argonne National Laboratory), Patricia A. Grubel (Los Alamos National Laboratory), Rinku K. Gupta (Argonne National Laboratory), and Gregory R. Watson (Oak Ridge National Laboratory)

This page provides detailed information specific to the tutorial event above. Expect updates to this page up to, and perhaps shortly after, the date of the tutorial. Pages for other tutorial events can be accessed from the main page of this site.

Quick Links

- Program Page (SC21 Website)
- Presentation Slides (FigShare)
- Hands-On Code Repository (GitHub)

On this Page

- Description
- Agenda
- Presentation Slides
- Hands-On Exercises
- Stay in Touch
- Resources from Presentations
- Requested Citation
- Acknowledgements

Description

Computational science and engineering (CSE) is in the midst of an extremely challenging period created by the confluence of disruptive changes in computing architectures, demand for greater scientific reproducibility, sustainability, and quality, and new opportunities for greatly improved simulation capabilities, especially through coupling physics and scales. These challenges demand increased investments in scientific software development and improved practices. Focusing on improved developer

Explaining Slide 2

- Slide 2 in all of our presentations contains the license, citation, and acknowledgements for the tutorial
- (Software) best practice to make your license and preferred citation(s) easily findable
- Sponsor acknowledgements rarely hurt!

License, Citation and Acknowledgements

License and Citation

- This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).
- **The requested citation the overall tutorial is:** Anshu Dubey, David E. Bernholdt, Better Scientific Software tutorial, in ISC High Performance, Hamburg, Germany and online, 2024

Acknowledgements

- Material included in these presentation is derived from work supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research (ASCR), and by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.
- This work was performed in part at the Argonne National Laboratory, which is managed by UChicago Argonne, LLC for the U.S. Department of Energy under Contract No. DE-AC02-06CH11357.
- This work was performed in part at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

We Want to Interact with You!

- We find these tutorials most interesting and informative (for everyone) if you ask questions and share experiences!
  - We learn too!
- Please raise your hand at any time to ask a question
- After the tutorial email us at bssw-tutorial@lists.mcs.anl.gov
  - With questions or feedback
  - The list moderator will allow your messages to be posted
- Refer to bssw-tutorial.github.io page for all tutorial materials

Agenda

The agenda is also available on the tutorial web page. Visit https://bssw-tutorial.github.io and click on the link for today's tutorial

Time (CEST)	Title
1:30 PM	Introduction
1:35 PM	Improving Reproducibility Through Better Software Practices
2:15 PM	Software Testing and Verification
3:00 PM	*Afternoon break*
3:30 PM	Managing Computational Experiments
4:30 PM	Reproducibility of Workflows
5:00 PM	*Adjourn*