



Better Software for Reproducible Science

David E. Bernholdt, Patricia A. Grubel,
David M. Rogers, and Gregory Watson

The International Conference for High-
Performance Computing, Networking, Storage,
and Analysis (SC23) conference

Final slides, hands-on activities,
last-minute updates, etc. at:
<https://bssw-tutorial.github.io/>
and click the link for today's tutorial




See slide 2 for
license details



License, Citation and Acknowledgements

License and Citation

- This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0). 
- **The requested citation the overall tutorial is:** David E. Bernholdt, Patricia A. Grubel, David M. Rogers, and Gregory R. Watson, Better Software for Reproducible Science tutorial, in The International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC23), Denver, Colorado, 2023. DOI: [10.6084/m9.figshare.24226105](https://doi.org/10.6084/m9.figshare.24226105).
- 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 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

- David E. Bernholdt, ORNL
- Patricia A. Grubel, LANL
- David M. Rogers, ORNL
- Gregory Watson, ORNL



David B
he/him



Patricia
she/her



David R
he/him



Gregory W
he/him

- Members of the IDEAS Productivity Project: <http://ideas-productivity.org>

The IDEAS-ECP team works with the ECP community, and beyond, to improve developer productivity and software sustainability as key aspects of increasing overall scientific productivity

- 1 Customize and curate methodologies**
- Target scientific software productivity and sustainability
 - Use workflow for best practices content development

- 2 Incrementally and iteratively improve software practices**
- Determine high-priority topics for improvement and track progress
 - *Productivity and Sustainability Improvement Planning (PSIP)*



- 3 Establish software communities**
- Determine community policies to improve software quality and compatibility
 - Create Software Development Kits (SDKs) to facilitate the combined use of complementary libraries and tools

- 4 Engage in community outreach**
- Broad community partnerships
 - Collaboration with computing facilities
 - Webinars, tutorials, events
 - *WhatIs* and *HowTo* docs
 - Better Scientific Software site (<https://bssw.io>)

For more about our work see this preprint:
<https://doi.org/10.48550/arXiv.2311.02010>

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 IDEAS and BSSw

- IDEAS Productivity mailing list: <http://eepurl.com/cQCyJ5>
 - Announcements of IDEAS-organized events
 - Best Practices for HPC Software Developers webinar series
 - Strategies for Working Remotely panel series
 - Software-focused events at major scientific meetings (e.g., SIAM, ISC, SC, etc.)
 - Typically 2-3 messages per month
- 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 of 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: David E. Bernholdt, Anshu Dubey, Patricia A. Grubel, Rinku K. Gupta, and Gregory R. Watson, Better Scientific Software tutorial, in the International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC21), St. Louis, MO, USA and online, 2021. DOI: [10.6084/m9.figshare.16556628](#)
- Individual modules may be cited as *Speaker, Module Title*, in Better Scientific Software tutorial...

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 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.
- 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 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.



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!
- 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

You may also be interested in these other software-related events at SC23:
<https://bssw.io/events/sc23-software-related-events>

(link is also on tutorial web page)

Agenda (1/2)

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 (MST)	Title	Presenter
8:30 AM	Introduction	Patricia A. Grubel (LANL)
8:40 AM	Motivation and Overview of Best Practices in HPC Software Development	Patricia A. Grubel (LANL)
9:05 AM	Scientific Software Design	David M. Rogers (ORNL)
9:30 AM	Refactoring Scientific Software	David M. Rogers (ORNL)
10:00 AM	<i>Morning break</i>	
10:30 AM	Software Packaging	David M. Rogers (ORNL)
11:00 AM	Collaborative Software Development	Patricia A. Grubel (LANL)
12:00 PM	<i>Lunch break</i>	

Agenda (2/2)

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 (MST)	Title	Presenter
12:00 PM	<i>Lunch break</i>	
1:00 PM	Improving Reproducibility Through Better Software Practices	Gregory R. Watson (ORNL)
2:15 PM	Reproducibility of Workflows	Gregory R. Watson (ORNL)
2:45 PM	Software Testing and Verification	David E. Bernholdt (ORNL)
3:00 PM	<i>Afternoon break</i>	
3:30 PM	Software Testing and Verification	David E. Bernholdt (ORNL)
3:45 PM	Lab Notebooks for Computational Mathematics, Sciences, & Engineering	David E. Bernholdt (ORNL)
4:45 PM	Summary	David E. Bernholdt (ORNL)
5:00 PM	<i>Adjourn</i>	