

reproducibility@xsede

An XSEDE14 Workshop

Reporting Back to Our Colleagues

Doug James, Carlos Rosales, Nancy Wilkins-Diehr

First, a sincere thanks to...

Allen, Gabrielle (E); Bailey, David (W); Barba, Lorena (EIC); Borwein, Jonathan (W); **Cazes, John** (E); Cho, Kym Won (E); **Colbry, Dirk** (EIWF); Corcho, Oscar (W); Deelman, Ewa (W); Dietze, Michael (W); **Fahey, Mark** (EIWF); Fonseca, Jim (E); Gilbert, Benjamin (W); Harkes, Jan (W); Hovig, Elvind (C); Hwang, Lorraine (E); **James, Doug** (EIWCF); Katz, Dan (EC); Kay, Sophie (C); Keele, Seth (W); **Koesterke, Lars** (E); **Kooper, Rob** (EIW); Kumar, Praveen (W); Lee, Jong (W); Lee, Kyo (EIF); **Lindsey, Susan** (EF); Linke, Erika (W); Loewe, Laurence (EWF); Marciano, Richard (W); **Marini, Luigi** (W); Mattman, Chris (W); Mattson, Dave (W); McHenry, Kenton (W); **McLay, Robert** (W); Miguez, Sheila (W); Minsker, Barbara (W); **Patel, Pragnesh** (EC); Perez-Hernandez, Maria (W); Pouchard, Line (E); **Rosales, Carlos** (ECF); **Roskies, Ralph** (EIF); Ryan, Dan (W); **Rynge, Mats** (W); Sandve, Geir (C); Santana-Perez, Idafen (W); Satyanarayanan, Mahadev (W); **Seidel, Ed** (E); Shi, Justin (EIWF); Silva, Rafael F. (EIWF); **Singer-Villalobos, Faith** (E); Skinner, David (C); St. Clair, Gloriana (W); Stodden, Victoria (EIWCF); Suriarachchi, Isuru (E); Webster, Keith (W); **Wilkins-Diehr, Nancy** (EIWCF)

Key: **Red** represents XSESE Staff or Campus Champion at time of event
(E): Event participant; (I): Invited presentation; (W): Written submission;
(C): Committee; (F): Final Report contribution

Outline

Brief overview of the workshop: Doug

Independent takeaways: Doug, Carlos, Nancy

Open discussion

reproducibility @ XSEDE: An XSEDE14 Workshop

Monday, July 14, 2014 - Atlanta, GA

Reproducibility

Submissions
Agenda
Final Report

Related Links:

XSEDE Home
XSEDE14 Annual Conference
Yale 2009 Roundtable
HPCWire Feature

Organizing Committee

reproducibility@XSEDE: An XSEDE14 Workshop

www.xsede.org/reproducibility

Format

~~Tutorial~~

~~Tech Track~~

Arm Wrestling



"Lt. j.g. Mary Robinson arm wrestles with a 6th grader from Shimoda Elementary School in Shimoda, Japan". U.S. Navy photo by Chief Information Systems Technician Kufere Usanga (2009). <http://commons.wikimedia.org>.

"...an interactive, open-ended, discussion-oriented agenda..."

A discussion about what?

"...focused on...large-scale computational science"

"...a response to the challenges issued in the Yale 2009 [Data and Code Sharing Roundtable] declaration"

"We hope to help promote a culture of reproducibility..."



"Lt. j.g. Mary Robinson arm wrestles with a 6th grader from Shimoda Elementary School in Shimoda, Japan". U.S. Navy photo by Chief Information Systems Technician Kufere Usanga (2009). <http://commons.wikimedia.org>.

Schedule

- Four 90-minute sessions

Session 0: Setting the Tone (Victoria Stodden)
Session 1: Where Do We Stand? (Dirk Colbry)
Session 2: What Can We Learn? (Lorena Barba)
Session 3: What Comes Next? (Nancy Wilkins-Diehr)

- Lightning Talks
 - Very loosely correlated with session themes
 - A professional courtesy: protected time for authors
 - Did not necessarily drive the discussion

reproducibility @ XSEDE: An XSEDE14 Workshop

Monday, July 14, 2014 - Atlanta, GA

Reproducibility

Submissions
Agenda
Final Report

Related Links:

XSEDE Home
XSEDE14 Annual Conference
Yale 2009 Roundtable
HPCWire Feature


Organizing Committee

reproducibility@XSEDE: An XSEDE14 Workshop

www.xsede.org/reproducibility

reproducibility @ XSEDE: An XSEDE14 Workshop

Monday, July 14, 2014 - Atlanta, GA

 Reproducibility

[Submissions](#)
[Agenda](#)
[Final Report](#)

Related Links:

[XSEDE Home](#)
[XSEDE14 Annual Conference](#)
[Yale 2009 Roundtable](#)
[HPCWire Feature](#)

Organizing Committee

Lorena A. Barba

George Washington University

Eivind Hovig

University of Oslo

Doug James (chair)

Texas Advanced Computing Center

Daniel S. Katz

University of Chicago & Argonne National Laboratory

Sophie Kay

University of Oxford

reproducibility@XSEDE: An XSEDE14 Workshop

Submissions

David Bailey and Jonathan Borwein

Fooling the masses : Reproducibility in high-performance computing

Dirk Colbry

Managing Advanced Computational Resources to Encourage Best Practices for Developing Repeatable Scientific Software

Mark Fahey and Robert McLay

Reproducibility Responsibilities in the HPC Arena

Doug James

Modeling Reproducibility for Reluctant Believers: A Parable in Bash

Laurence Loewe and Seth A. Keel

BEST Names for Semantic Units to Support Reproducible Modeling

Kenton McHenry, Jong Lee, Michael Dietze, Praveen Kumar, Barbara Minsker, Richard Marciano, Luigi Marini, Rob Kooper, Dave Mattson

DIBBs Brown Dog, PaaS for SaaS for PaaS

Chris A. Mattmann

8:30 - 10:00

Session 0: Reproducibility in Large-Scale Computing: Setting the Tone (Victoria Stodden)

Presentation Slides

- Reproducibility in Computational Science (Victoria Stodden) "I have yet to meet a scientist without some intuitive understanding that reproducibility is a desirable component of scientific research. However, opinions differ widely on the scope and seriousness of the reproducibility issue, and on which actions or solutions are appropriate. In this talk I will outline some of the history of the reproducibility discussion in the scientific community, from the Yale 2009 Roundtable on Data and Code Sharing to more recent efforts. I will use this narrative to motivate one frame for addressing the why, what, and how of reproducible computational science."
- Guided Discussion

10:00 - 10:30

Break

10:30 - 12:00

Session 1: Reproducibility in Large-Scale Computing: Where Do We Stand? (Dirk Colbry)

- Five-minute lightning talks: Mark Fahey, Rob Kooper, Justin Shi, Nancy Wilkins-Diehr
- Guided Discussion

Final Report available
at [workshop website](#)
and elsewhere

Standing Together
for
Reproducibility in Large-Scale Computing

Report on reproducibility@XSEDE
An XSEDE14 Workshop
July 14, 2014
Atlanta, GA

Developed collaboratively by the reproducibility@XSEDE workshop participants¹

Principal Editors:
Doug James, Nancy Wilkins-Diehr, Victoria Stodden, Dirk Colbry, and Carlos Rosales

Finalized 17 Dec 2014

Abstract. *This is the final report on reproducibility@xsede, a one-day workshop held in conjunction with XSEDE14, the annual conference of the Extreme Science and Engineering Discovery Environment (XSEDE). The workshop's discussion-oriented agenda focused on reproducibility in large-scale computational research. Two important themes capture the spirit of the workshop submissions and discussions: (1) organizational stakeholders, especially supercomputer centers, are in a unique position to promote, enable, and support reproducible research; and (2) individual researchers should conduct each experiment as though someone will replicate that experiment. Participants documented numerous issues, questions, technologies, practices, and potentially promising initiatives emerging from the discussion, but also highlighted four areas of particular interest to XSEDE: (1) documentation and training that promotes reproducible research; (2) system-level tools that provide build- and run-time information at the level of the individual job; (3) the need to model best practices in research collaborations involving XSEDE staff; and (4) continued work on gateways and related technologies. In addition, an intriguing question emerged from the day's interactions: would there be value in establishing an annual award for excellence in reproducible research?*

Themes more than conclusions

"Two important themes capture the spirit of the workshop submissions and discussions:

- (1) organizational stakeholders, especially supercomputer centers, are in a unique position to promote, enable, and support reproducible research... [Fahey]*
- (2) individual researchers should conduct each experiment as though someone will replicate that experiment [Barba]."*



*“If everyone's thinking alike,
then nobody's thinking
(and everybody's wrong).”*

-- Attribution and wording varies

Benjamin Franklin 1767 (David Martin).
Project Gutenberg (from Wikipedia Commons).
Copyright expired; Image is in the public domain
http://commons.wikimedia.org/wiki/File:Benjamin_Franklin_1767.jpg.

Doug's Takeaways*

*Others' mileage WILL vary.
In fact I confess to being deliberately provocative.

Where Do I Stand?*

“No, thank you...”

- Run your code
- Duplicate your environment
- Bitwise reproducibility

“Yes, please...”

- Implement your algorithm
- Portable results
- Understand the differences

reproducibility > re-enactment

***I speak only for myself!**

Definitions can get in the way

- Narrow definitions are unnecessary
 - Bitwise reproducibility, virtual environments, ...
- Broad definitions are not helpful
 - Reproducibility = everything
 - Reproducibility = whatever I'm working on

My own Goldilocks perspective*

- Reproducibility is the ability to **revisit** an experiment **to the degree necessary** to assess its correctness and impact
 - "revisit" might mean repeat, re-implement, review, ...
- It's about establishing confidence
 - visibility, provenance, audit trails, record keeping, automation: knowing (and communicating) how you got where you are
 - the experiment, not the software

Reproducibility makes my life easier

- Not something you bolt on at the end so somebody else can duplicate your results.
- Propels your own work forward in ways that make you more efficient, more confident, and more effective.
- Our scripts, repositories, and processes...
 - allow us to resolve issues, recognize patterns, and progress more quickly than would otherwise have been possible.
 - are templates, samples, and demos that we can leverage next time around
- Sandve et al: "you should at least be able to reproduce the results yourself"

Model Good Behavior

- Collaborations (including ECSS and tickets)
 - "A reproducible problem is a wonderful thing"
 - "Give me [or 'here is'] a small reproducer"
 - Change management, automation, provenance
 - A paragraph or two in work plans and final reports
- Training and Mentoring
 - I'd rather not see a course called "Reproducibility"
 - I prefer ad hoc, just-in-time demos leading to "I want a piece of that action" moments
 - We don't need to re-invent wheels (e.g. Software Carpentry)

Essential References

- Sandve, Nekrutenko, Taylor, and Hovig (2013) [Ten Simple Rules for Reproducible Computational Research.](#)
- McConnell (1993, revised 2004), [Code Complete: A Practical Handbook of Software Construction, Second Edition.](#)



*“If everyone's thinking alike,
then nobody's thinking
(and everybody's wrong).”*

-- Attribution and wording varies

Benjamin Franklin 1767 (David Martin).

Project Gutenberg (from Wikipedia Commons).

Copyright expired; Image is in the public domain

[http://commons.wikimedia.org/wiki/File:Benjamin](http://commons.wikimedia.org/wiki/File:Benjamin_Franklin_1767.jpg)

[_Franklin_1767.jpg](http://commons.wikimedia.org/wiki/File:Benjamin_Franklin_1767.jpg).

Carlos' Takeaways

Before the Workshop

- I had never considered this issue in detail
- Opinion: Research must be reproducible –
How naïve!!!
- But what does this mean???

What does reproducibility mean?

- Many interpretations in the workshop
- Lots of idealistic views – bitwise reproducibility at an indeterminate time in the future – **NOT PRACTICAL**
- My own interpretation after the workshop:

Perform and document research as if somebody else were going to reproduce it

That is a very loose definition...

- YES. And it is the way we have been doing scientific research for more than a century.
- Your peers must have confidence in
 - The honesty of your research
 - The completeness of your report
 - Their ability to reproduce your investigation
- This is still quite a high bar!

Do we have the tools to do this?

- Yes and no...
- Many available tools
 - Open Source and Creative Commons licensing
 - Free open source repositories
 - Ability to acquire DOI for code and documentation
 - Environment capture and logging (XALT)
- BUT many obstacles too
 - Universities attitude towards Open Source
 - Deployment of system tools is limited
 - Information about resources not easy to find

Making it work

- Willingness to do it!
 - It is its own reward
 - Encouragement / formal recognition
- Wider dissemination of resource location
 - Document resources for XSEDE users
- Wider deployment of logging tools
 - All XSEDE resources using XALT or equivalent
- Involve universities in the process
 - Easier access to resources
 - Less red tape for Open Source / Open Commons

Summary

- I was surprised by the level of interest and the energetic discussions
- I realized my ideas were naïve and incomplete
- There are tools out there to support reproducible work
 - XSEDE can make it easier for researchers
 - But the responsibility falls on the individual



*“If everyone's thinking alike,
then nobody's thinking
(and everybody's wrong).”*

-- Attribution and wording varies

Benjamin Franklin 1767 (David Martin).

Project Gutenberg (from Wikipedia Commons).

Copyright expired; Image is in the public domain

[http://commons.wikimedia.org/wiki/File:Benjamin](http://commons.wikimedia.org/wiki/File:Benjamin_Franklin_1767.jpg)

[_Franklin_1767.jpg](http://commons.wikimedia.org/wiki/File:Benjamin_Franklin_1767.jpg).

Nancy's Takeaways

April 21, 2015

Reproducibility in Large-Scale Computing: What Comes Next?

XSEDE

Extreme Science and Engineering
Discovery Environment




Three goals for the workshop

- Highlight the issue
 - Begin a discussion within XSEDE
 - Discuss the unique challenges around reproducibility at scale
- Understand what's needed to create a culture of reproducibility
 - How can XSEDE support stakeholders and users in these endeavors
- Communicate ideas from the workshop to XSEDE leadership

Final session focused on what XSEDE can do

Mozilla Etherpad used to capture ideas

Version 7693 Saved September 30, 2014

 Return to pad

Authors: Line Pouchard -Purdue, Victoria, Mark Fahey, Susan Lindsey, Nancy Wilkins-Diehr, SDSC, Doug James, Lorena Barba, GW, Laurence
Loewe, Daniel S. Katz, NSF, Rafael F. Silva - USC/ISI + 2 unnamed authors ()

Reproducibility in Large-Scale Computing: What Comes Next?

- **repeat an experiment to the degree necessary to assess the validity and importance of the claimed results**
- **ability to independently confirm a scientific hypothesis and discover/resolve differences in implementation**
- Our goals for this day were twofold
 - Call attention to this important topic in the XSEDE space. I've worked at a center for 20 years and have never attended a talk on reproducibility. Some work is implicit in the way computational science is done - changing platforms all the time, sys admins check results after upgrades, users check results when changing platforms or adding new science code and after upgrades (sometimes). What is a more thoughtful, systematic approach?
 - **Promote a culture of reproducibility (Doug's opening)**
 - Understand what's different about reproducibility for XSEDE (i.e., at large scale) vs for computational and data science in general
 - Develop recommendations, for ourselves and for our user community
 - Next steps, priorities, initiatives
 - Final report will include priorities and initiatives for organizations and communities
 - Perhaps serve as a model for others

Yale 2009 long term goals

1. Version control system for data.
2. Publish code and testing suites.
3. Tools to cite code and data.
4. Tools for download tracking
5. Mark reproducible published docs as such.
6. Require authors to use standard ontologies.
7. Universities archive research results
8. Clarify ownership issues.
9. Develop deeper communities that maintain code and data.

Write your name below (the color will help identify your contributions)



XSEDE

Here's what we came up with

- Develop and structure documentation and training materials in ways that encourage and enable reproducible research
 - When possible, promote and leverage existing resources, e.g. training activities like those pioneered by the Software Carpentry Initiative and literature like Ten Simple Rules for Reproducible Computational Research
- Evaluate XALT as a mechanism to collect and record detailed build- and run-time software usage data (compilers, libraries, version information) for users' jobs
 - Include SPs and Champion sites
- Use ECSS to model best practices
 - Conducting research activities in a reproducible way
 - Mentoring research groups in process-related matters
 - Possibly addressing reproducibility in work plans and final reports
- Include reproducibility as a focus when working with science gateways and workflows
 - Both are well-suited to deliver user-friendly workflows, automated processes, and repeatable results
- Investigate establishment an annual award for excellence in reproducible research



*“If everyone's thinking alike,
then nobody's thinking
(and everybody's wrong).”*

-- Attribution and wording varies

Benjamin Franklin 1767 (David Martin).

Project Gutenberg (from Wikipedia Commons).

Copyright expired; Image is in the public domain

[http://commons.wikimedia.org/wiki/File:Benjamin](http://commons.wikimedia.org/wiki/File:Benjamin_Franklin_1767.jpg)

[_Franklin_1767.jpg](http://commons.wikimedia.org/wiki/File:Benjamin_Franklin_1767.jpg).

Open Discussion

Time for your input

We would like to hear from participants on this call

- What do you find challenging about reproducibility?
- What could XSEDE do that would make that easier?
- How can we help users conduct their research *as though* someone will replicate their work?