



# Research Computing in Child Development Research

Penn State Child Study Center (CSC)

Rick Gilmore

2020-02-25 08:00:12

# Agenda

- Why are we here
- Research computing in developmental research
- Where are we now
- Where do we want to go
- Preview of strategic plan elements

Why are we here

- Open Data & Developmental Science (ODDS) initiative
  - R Bootcamps, [2017](#), [2018](#), [2019](#) (with M. Hallquist)
  - Grad [course\(s\)](#) on reproducible research practices
  - Policies and best practices (e.g., Gilmore, Cole, et. al., in press, *Child Development Perspectives*; [SRCD Policy](#))
  - [Databrary.org](#); [Play & Learning Across a Year \(PLAY\) project](#)
  - Local hosting of [ABCD](#) database (M. Hallquist)

- Data science initiatives
  - [Institute for Computational & Data Science \(ICDS\)](#)
  - [Center for Social Data Analytics \(SoDA\)](#)
  - Data Science Community meetings

- New faces at Old Main
  - New Senior VP for Research ([Lora Weiss](#))
  - New Associate CIO for Research Computing ([Greg Madden](#))
  - National search for new [CIO/VP IT](#)
- Faculty governance of research computing
  - [Research Computing & Cyberinfrastructure \(RCCI\)](#) groups
- Research computing services [web site](#)
- University-wide strategic plan update this spring

# Research computing in developmental research

- How we *gather* data
  - Instruments (scanners, EEG, physio, eye tracking, smartphones/tablets)
  - Digital video/audio
  - Computer-based tasks, surveys
- Where/how we *store* data
- Where/how we *analyze* data
- How we *protect* data
- Where/how/when/with whom we *share* data & materials



*"How can our faculty and students be more productive?"*

*"What resources (people, technologies, and expertise) will enable Penn State's developmental community to expand its impact and reach?"*

*"What barriers (policies, technologies, and expertise) limit or slow progress?"*

*"How do emerging trends around sharing, transparency, and openness affect our work?"*

**Where we are now...**

# The “Rumsfeldian”



- What are we confident we know? (known-knowns)
- What are we confident we don't know? (known-unknowns)
- What might we have overlooked, neglected, or forgotten? (unknown-knowns)

# What's working well?

...for you with research computing...

- How we *gather* data
- Where/how we *store* data
- Where/how we *analyze* data
- Where/how/when/with whom we *share* data (and materials)
- technologies, expertise, policies

**What's not working well?**

Where do we want to go



- What aren't we doing but could be?
- What must we continue? What might be streamlined or phased-out?
- What are the highest priority/highest impact areas where we need to improve or build capacity?

# Preview of strategic plan elements

# Researcher Support

- We will ensure that researchers have **access to national-caliber advanced computational services for traditional high performance computing** and high throughput computing workloads.
- We will ensure that **researchers working with emerging technologies** (e.g. artificial intelligence, machine learning, immersive technologies, etc.) **have access to the tools and expertise** they need to accomplish their research.

- We will ensure that researchers have access to resources and expertise to accomplish their goals whether they are working on local-scale computational clusters, university-scale computational clusters, national-scale computational clusters, or cloud-native computational environments.
- We will ensure that researchers have **access to modern software for administering** their research.

- We will ensure that researchers have **access to appropriate resources for providing public access to their research**, whether through website tools or university-scale analytical tools.
- We will ensure that researchers **have the tools to collaborate** effectively across the university, the nation, and the world.

# Research Data Support

- We will partner with the research-data-related offices across the university to **put in place a holistic approach to managing research data** from data acquisition through data archiving or disposal.
- Through a combination of network technologies and data transfer applications, we will **ensure that large datasets can be efficiently moved** around the university, and to partners elsewhere, as needed.

- We will partner with an appropriate subset of the research-data-related offices across the university to **ensure that the university can properly secure any research data** subject to access restrictions.
- We will partner with an appropriate subset of the research-data-related offices across the university to ensure that the university can properly **comply with all public access to data requirements**.

**What's missing?**



**What should be the most important priorities and why?**

# Resources

# Software

This talk was produced on 2020-02-25 in [RStudio](#) using R Markdown. The code and materials used to generate the slides may be found at <https://github.com/psu-psychology/open-data-and-developmental-science-ODDS/>. Information about the R Session that produced the code is as follows:

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats  graphics  grDevices  utils  datasets
## [6] methods  base
```