Agenda

• Introductory comments
• Presentations: Andi Rauber, others?
• Conclusions and Next Steps
RIG Goals

1. Theme: Where does reproducibility fit in the RDA structure? Can we leverage the work of other IGs and WGs?

2. What are tools that support reproducibility? Can we collate a list? Find gaps?

3. Use cases for reproducibility research. Exemplars.

(4.) Can we match tools and use cases?
Update and Recap

Previous meetings:

• RDA-4: 1st meeting; lots of interest and lively discussion

• RDA-5: joint session with Provenance WG

• RDA-6: Google doc: http://bit.ly/2cCE2Q1 (or https://docs.google.com/document/d/18ptKKJQJLOC4B71Mcd9mATyYxOTQYMza0w2vpg-GSp7E/edit )
Parsing Reproducibility

“Empirical Reproducibility”

“Statistical Reproducibility”

“Computational Reproducibility”

V. Stodden, IMS Bulletin (2013)
Computational Reproducibility

Traditionally two branches to the scientific method:

• Branch 1 (deductive): mathematics, formal logic,

• Branch 2 (empirical): statistical analysis of controlled experiments.

Now, new branches due to technological changes?

• Branch 3,4? (computational): large scale simulations / data driven computational science.

CLAIM: computation presents only a potential third/fourth branch of the scientific method (Donoho et al 2009).
Infrastructure Responses

Tools and software to enhance reproducibility and disseminate the scholarly record:

Dissemination Platforms

- ResearchCompendia.org
- MLOSS.org
- IPOL
- Madagascar
- Open Science Framework
- thedatahub.org
- nanoHUB.org
- RunMyCode.org

Workflow Tracking and Research Environments

- Vistrails
- Kepler
- CDE
- Jupyter
- torch.ch
- Galaxy
- GenePattern
- Sumatra
- Taverna
- DataCenterHub
- Pegasus
- Kurator
- RCloud

Embedded Publishing

- Verifiable Computational Research
- SOLE
- knitR
- Collage Authoring Environment
- SHARE
- Sweave
Three Principles for CI

1. *Supporting scientific norms*—not only should CI enable new discoveries, but it should also permit others to reproduce the computational findings, reuse and combine digital outputs such as datasets and code, and facilitate validation and comparisons to previous findings.

2. *Supporting best practices in science*—CI in support of science should embed and encourage best practices in scientific research and discovery.

3. *Taking a holistic approach to CI*—the complete end-to-end research pipeline should be considered to ensure interoperability and the effective implementation of 1 and 2.

Changes embedded in a social and political environment.

Exceptions: privacy, HIPAA, FERPA, other constraints on sharing.

Community Responses

Declarations and Documents:

- Yale Declaration 2009
- ICERM 2012
- XSEDE 2014
Really Reproducible Research

“Really Reproducible Research” (1992) inspired by Stanford Professor Jon Claerbout:

“The idea is: An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures.” David Donoho, 1998

Note the difference between: reproducing the computational steps and, replicating the experiments independently including data collection and software implementation. (Both required)
Querying the Scholarly Record

- Show a table of effect sizes and p-values in all phase-3 clinical trials for Melanoma published after 1994;

- Name all of the image denoising algorithms ever used to remove white noise from the famous “Barbara” image, with citations;

- List all of the classifiers applied to the famous acute lymphoblastic leukemia dataset, along with their type-1 and type-2 error rates;

- Create a unified dataset containing all published whole-genome sequences identified with mutation in the gene BRCA1;

- Randomly reassign treatment and control labels to cases in published clinical trial X and calculate effect size. Repeat many times and create a histogram of the effect sizes. Perform this for every clinical trial published in the year 2003 and list the trial name and histogram side by side.
Government Mandates

• OSTP 2013 Open Data and Open Access Executive Memorandum; Executive Order.

• “Public Access to Results of NSF-Funded Research”

• NOAA Data Management Plan, Data Sharing Plan

• NIST “Common Access Platform”

• …
Federal Agencies

Reliable Science: The Path to Robust Research Results

September 8, 2015

These days, much discussion about the reproducibility of scientific results seems driven by critiques of research in biomedicine and psychology. Most recently, an article in Science concluded that 60 percent of a collection of studies were not replicable. This result along with similar analyses of cancer research results have stimulated strong commentary. For example, the New York Times print edition headline about the Science article was “Psychology’s Fears Confirmed: Rechecked Studies Don’t Hold Up,” coverage that prompted a strong op-ed rebuttal titled, “Psychology Is Not in Crisis.”

Issues that arise with human subjects or with other complex living systems do not plague physical science to the same degree. However, the notion of measuring the same value of a physical quantity or the same behavior of a physical system in different laboratories at different times is central to our concept of a valid scientific result. Often the approach is simply to replicate an experiment, but rather to get at the same quantity via different paths. For example, we can measure the gravitational constant, G, with approaches ranging from a torsional pendulum to atom interferometry.

Two of the cornerstones of science advancement are rigor in designing and performing scientific research and the ability to reproduce biomedical research findings. The application of rigor ensures robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results. When a result can be reproduced by multiple scientists, it validates the original results and readiness to progress to the next phase of research. This is especially important for clinical trials in humans, which are built on studies that have demonstrated a particular effect or outcome.

In recent years, however, there has been a growing awareness of the need for rigorously designed published preclinical studies, to ensure that such studies can be reproduced. This webpage provides information about the efforts underway by NIH to enhance rigor and reproducibility in scientific research.
Journal Requirements

• Science: code data sharing since 2011.
• Nature: data sharing.
• AER: data and code access
• others

The Larger Community

1. **Production**: Crowdsourcing and public engagement in science primarily data collection/donation today, but open up pipeline:
   - access to “coherent” digital scholarly objects,
   - mechanism for ingesting/evaluating new findings,
   - addressing legal issues (use, re-use, privacy,…).

2. **Use**: “Evidence-based”-{policy, medicine, …}, decision making.
Open Questions

• Incentivizing changes toward the production and dissemination of reproducible research.

• Who funds and supports cyberinfrastructure? Who controls access and gateways?

• Who owns data, code, and research outputs? Working around and within blocks such as privacy, legal barriers, ..

• What are community standards around documentation, citation standards, best practices? Who enforces?
Empirical Reproducibility

Sorting Out the FACS: A Devil in the Details

William C. Hines, Ying Su, Irene Kuhn, Kornelia Polyak, and Mina J. Bissell

The reproduction of results is the cornerstone of science; yet, at times, reproducing the results of others can be a difficult challenge. Our two laboratories, one on the East and the other on the West Coast of the United States, decided to collaborate on a project of mutual interest—namely, the heterogeneity of the human breast. Despite using seemingly identical methods, reagents, and specimens, our two laboratories quite reproducibly were unable to replicate each other’s fluorescence-activated cell sorting (FACS) profiles of primary breast cells. Frustration of studying cells close to their context in vivo makes the exercise even more challenging.

Paired with in situ characterizations, FACS has emerged as the technology most suitable for distinguishing diversity among different cell populations in the mammary gland. Flow instruments have evolved from being able to detect only a few parameters to those now capable of measuring up to—and beyond—an astonishing 50 individual markers per cell (Cheung and Ul, 2011). As with any exponential increase in data complexity, breast reduction mammoplasties. Molecular analysis of separated fractions was to be performed in Boston (K.P.’s laboratory, Dana-Farber Cancer Institute, Harvard Medical School), whereas functional analysis of separated cell populations grown in 3D matrices was to take place in Berkeley (M.J.B.’s laboratory, Lawrence Berkeley National Lab, University of California, Berkeley). Both our laboratories have decades of experience and established protocols for isolating cells from primary normal breast tissues as well as the capabilities required for...
Statistical Reproducibility

- False discovery, p-hacking (Simonsohn 2012), file drawer problem, overuse and mis-use of p-values, lack of multiple testing adjustments.
- Low power, poor experimental design, nonrandom sampling,
- Data preparation, treatment of outliers, re-combination of datasets, insufficient reporting/tracking practices,
- inappropriate tests or models, model misspecification,
- Model robustness to parameter changes and data perturbations,
- Investigator bias toward previous findings; conflicts of interest.
- …
Background: Open Source Software

- Innovation: Open Licensing
  
  ➡ Software with licenses that communicate alternative terms of use to code developers, rather than the copyright default.

- Hundreds of open source software licenses:
  
  - GNU Public License (GPL)
  - (Modified) BSD License
  - MIT License
  - Apache 2.0 License

- ... see http://www.opensource.org/licenses/alphabetical
The Reproducible Research Standard

The Reproducible Research Standard (RRS) (Stodden, 2009)

- A suite of license recommendations for computational science:
  - Release media components (text, figures) under CC BY,
  - Release code components under Modified BSD or similar,
  - Release data to public domain or attach attribution license.
- Remove copyright’s barrier to reproducible research and,
- Realign the IP framework with longstanding scientific norms.
Research Compendia

Pilot project: improve understanding of reproducible computational science, trace sources of error

• link data/code to published claims, re-use,
• a guide to empirical researchers,
• certifies results,
• large scale validation of findings,
• stability, sensitivity checks.
Is “Huh?” a Universal Word? Conversational Infrastructure and the Convergent Evolution of Linguistic Items

Mark Dingemanse, Francisco Torreira, N. J. Enfield, Johan J. Bolhuis

Code and Data Abstract

A word like Huh?—used as a repair initiator when, for example, one has not clearly heard what someone just said—is found in roughly the same form and function in spoken languages across the globe. We investigate it in naturally occurring conversations in ten languages and present evidence and arguments for two distinct claims: that Huh? is universal, and that it is a word. In support of the first, we show that the similarities in form and function of this interjection across languages are much greater than expected by chance. In support of the second claim we show that it is a lexical, conventionalised form that has to be learnt, unlike grunts or emotional cries. We discuss possible reasons for the cross-linguistic similarity and propose an account in terms of convergent evolution. Huh? is a universal word not because it is innate but because it is shaped by selective pressures in an interactional environment that all languages share: that of other-initiated repair. Our proposal enhances evolutionary models of language change by suggesting that conversational infrastructure can drive the convergent cultural evolution of linguistic items.
The MIT License (MIT)

Copyright (c) 2013 Sheila Miguez, Victoria Stodden, Jennifer Seiler

Permission is hereby granted, free of charge, to any person obtaining a copy
of this software and associated documentation files (the "Software"), to
deal in the Software without restriction, including without limitation the
rights to use, copy, modify, merge, publish, distribute, sublicense, and/or
sell copies of the Software, and to permit persons to whom the Software is
furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in
all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING
FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS
IN THE SOFTWARE.