Reproducibility Notes

NCSA Reproducibility Group

November 10, 2017

Contents

1 Contributors

This document contains work from the NCSA Reproducibility group. The contributors are listed below.

- Matthew Krafczyk

- Adhithya Bhaskar

2 Introduction

These are my research notes while working at NCSA.
    Complete

3 Reproducibility Terms

I will be using the ACM terms for reproducibility [1]. In order of increasing strength or rigor they are:

- **Repeatability**: When a research team can get the same conclusions using the same experimental methods as previous

- **Replicability**: When a separate research team can get the same conclusions using the same experimental methods as previous

- **Reproducibility**: When a separate research team can get the same conclusions using a different experimental method as previous

There is some contention about these terms, I believe the biology people use different terms.   Needs Citation

3

4   Reproducibility Issues

Here I'm trying to collect all the different issues which exist with reproducibility. I'm using it as a space to make sense of these issues as well as collect the references which make the case.

Complete

4.1   Sociological Issues

- Scrutinization
- Competitive Advantage
- Licensing
- Monolithic codebases
- Dirty code
- **Some seem *strange* given the published article**

Complete

4.2   Software Issues

- Numerical error
- Transparent machine software/hardware failures
- Hard-coded machine topologies
- **Some of these are *unavoidable***

Complete

4.3   Hardware Issues

- Special hardware requirements
- Silent Failures
- **Some of these are *unavoidable***

Complete

4.3.1   Specialized Hardware

Here I'm detailing various specialized hardware I hear about which might hamper reproducibility efforts.

- Large Supercomputer (for large computing runs)
- GPUs (Usually not hard to get, but still an obstacle)
- FPGA
- Neuromorphic Chips [2]
- TPUs [3]

Complete

4.3.2 Handling specialized hardware

Results run on specialized hardware should still be replicable as long as an appropriate virtual machine or emulator is available for that architecture.

Complete

4.4 Compatibility Issues

DOES THIS SUBSECTION DESERVE ITS OWN SECTION??

Study compatibility is a serious problem which prevents the comparison of scientific studies when researchers do not agree on input formatting, output formatting, and general scientific workflow. Without this compatibility, major work must be done to allow apples to apples comparisons between scientific works.

To make matters worse, allowing scientists to steal a method from another scientist and use it in their IDE is currently impossible due to incompatibilities inherent in each application's assuredly divergent development lines. Thus solving the compatibility will be a great step forward towards allowing the three currently difficult workflows.

Complete

5 Reproducibility Scandals

Here, I'm going to collect together known scandals arrising from the lack of reproducibility.

5.1 Medicine

Medicine specific scandals

- 2004 GlaxoSmithKline Paxil uneffective [4]. Needs Citation

- 2011 Bayer could only reproduce 17 of 67 published studies [4]. Needs Citation

- 2012 Amgen could only reproduce 6 of 53 published cancer studies [4]. Needs Citation

- 2014 Tamiflu review showed only small decrease in some smptoms [4]. Needs Citation

- 2017 A doctor who developed a scale for measuring drug regime adherence in patients is extracting licensing fees from users of the scale. This makes reproducing work using said scale more difficult. [5]

5.2 Physics

Physics specific scandals.

- BICEP2 Gravitational Wave 'discovery' [4]. Needs Citation

Complete

5.3 Chemistry

Chemistry specific scandals.

- Gaussian software banning competitors from using their software

5.3.1 Gaussian

Gaussian Inc. actively bans competing researchers from using their code. This means if you are working on a tool which performs the same tasks as Gaussian and probably especially if you claim to do a better job than Gaussian, you are banned from using Gaussian forever. This means that said researchers cannot validate research published by other scientists who do use the Gaussian software. The website [6] goes into great detail about this ongoing crisis. They furthur reference journal standards produced by the Pure and Applied Chemistry journal [7], [8] which make mention of the requirement that sufficient technical details be published to permit the replication of any published article's results. Gaussian banning scientists directly contradicts this stance by the journal.

Complete

6 Causation of Reproducibility problems

This section summarizes why reproducibility is a problem in various fields and not other.   Complete

7 Numerical Precision Issues

This section summarizes issues related to numerical precisions. Whether it's an algorithm touted to be more precise, or a tool to optimize a program it will be found here.

In calculations which are sensitive to numerical precision, small errors may lead to large numerical problems such as taking the wrong branch in a conditional statement [9].

Complete

7.1 IEEE floating point arithmetic

Here, I review some properties of the IEEE floating point standard. I may later add some information about some fixed-point or alternative floating point implementations.

The IEEE 754 floating point standard [10], [11] defines operation and exception details for the floating point number.

Each floating point number consists of a 'mantissa' and an 'exponent'. The mantissa (m) is a signed integer of precision p and the exponent (e) is a signed integer with precision e. Each floating point number is also defined by a radix(b) which can either be 2(binary) or 10(decimal). Each floating point number then has a value $m \cdot b^e$. Table 1 defines these parameters for each defined binary type.

Table 1: Floating point variable parameters for each binary (radix 2) defined type. All taken from [10].

bit-width	32	64	128
specification name	binary32	binary64	binary128
colloquial name	`float`	`double`	`quad`
precision	24	53	113
exponent precision	8	11	15
equivalent decimal precision	7 digits	16 digits	34 digits

7.1.1 Normalization

A common way of expressing a floating point number is like this: $\pm \beta^e \times .d_1 d_2 \ldots d_t$ with $d_1 \neq 0$ for *normalized* numbers. when y is less than β^{emin}, d_1 cannot be non-zero and so we have sub-normal numbers [12].

7.1.2 Rounding modes

From libc and ieee [10], [13], we have four rounding modes. [13] says that fpu calculations occur at higher precision, and then the value is stored in the specified data width where rounding occurs. I haven't found the statement in [10] which says this.

- Round to nearest (default)

- Round to $+\infty$

- Round to $-\infty$

- Round to 0

7.1.3 Machine Epsilon

From lemma 2.1 on page 37 of [12], The spacing between a normalized floating point number x and an adjacent normalized floating point number is at least $\beta^{-1}\epsilon|x|$ and at most $\epsilon|x|$ [12]. This also means that the largest error between an infinitely precise number, and it's nearest neighbor is $\epsilon/2 = \beta^{-(p-1)}/2$. This value is known as the machine epsilon.

Another definition: Machine epsilon is defined as the smallest number that, when added to one, yields a result different from one. This is $b^{-(p-1)}$ and $\epsilon/2$ for the round-to-nearest procedure.

7.1.4 Known numerical anomalies

Bailey goes through a great discussion of known numerical anomalies here [14]. I'm going to talk about a couple of them.

- Language-based difficulties

    - code constant translation to lower precision types than those on the rhs of equations.

    - Expression evaluation to the precision of only the operand types and not the type on the rhs of the equation.

- Avoidable Numerical Anomalies

    - Conditional branches with floating-point comparison which may fail to hold

    - error accumulation from summation of small numbers

- Unavoidable Numerical Anomalies

    - Large equations or systems of equations are demanding numerically.

7.1.5 Reproducibility Attributes

This section discusses the reproducibility attributes present in the IEEE 754 standard. [10]
    Complete

7.2 Scientific Problems needing special precision

A number of problems need special scientific precision. David Bailey did a great job laying this out[4], [9]. I reproduce his table from [4] as Table 2 to give an idea of the gamut.

Table 2: Typical bit precision necessary for various scientific fields

Problem Type	Number of bits(?) necessary
Planetary orbit calculations	32
Climate modeling	32
Optimization problems in biology and other fields	32
Schrodinger solutions for lithium and helium atoms	32
Scattering amplitudes of fundamental particles	32
Discrete dynamical systems	32
Supernova simulations	32-64
Coulomb n-body atomic system simulations	32-120
Electromagnetic scattering theory	32-100
The Taylor algorithm for ODEs	100-600
Ising integrals from mathematical physics	100-1,000
Problems in experimental mathematics	100-50,000

MINOS Needs Citation software for biochemistry contains precision problems [4].

7.3   Solutions to reduction error issues

Two techniques which can be used to solve reduction error are to either enforce an ordering to the summation. This can be done by creating a unique id for each element of the sum, and guarenteeing that these ids are identical each run and making sure that the summation sums the id'd elements the same way each time. Another is to use the technique outlined here: [15].

There are some techniques to test whether your program has numerical precision difficulties. You can for instance, change the rounding mode. If the program's result varies wildly, you know there is a precision problem, but it does nothing to clear this up [14]. You can add an adjustable 'fuzz' to a computation to see whether this affects the results offered. This can prove numerical instability [14].

7.4   Software implementing arbitrary precision floating point

Here I collect packages which implement arbitrary precision floating point arithmetic in various ways. These packages were discovered in [9].

- ARPREC (Arbitrary, pre-set precision level) Needs Citation

- DDFUN (double double precision) Needs Citation

- FMLIB/FMZM90 (low-level multiple-precision libarary) Needs Citation

- GMP (multiple-precision library) Needs Citation

- Predicates (Specialized package for computational geometry) Needs Citation

- QD (double double and quad double arithmetic. Faster than arbitrary precision libraries when only double double or quad double precision is sufficient) Needs Citation

- MPFR/MPFR++ (multi-precision library with exact rounding Needs Citation

- MPFUN90 (ARPREC equivalent entirely in fortran90) [16]

- VPA (arbitrary precision library for fortran) Needs Citation

From [4], we have:

- CORVETTE Needs Citation

Using high-precision software increases runtimes. double-double Takes 5 times longer than 64-bit arithmetic, 25 times longer with quad-double, and 50 times for 100-digit arithmetic. Computational cost scales as p^2 up to about 1000 digits then costs $p \log p$ if you're using FFT-based multiplication. [9] (Needs Citation for FFFT-based multiplication)

7.5   Program Precision Optimizers

These tools analyze a program's operation in some way and produce recommendations about data types or run conditions which will preserve or increase a program's numerical precision while also improving or largely preserving a programs performance.

- BLAME Analysis [17]

- PRECIMONIOUS Needs Citation

A few examples of applications which have benefitted from these techniques or detailed studies include climate modelling, supernova simulations, coulomb n-body simulations, studies of the fine-structure constant, E-M scattering theory, vortex sheet roll-up (fluid dynamics) simulations, computational geometry, compuational number theory, experimental math (including the discovery of the BBP π formula, quantum field theory (Using PSLQ computations (More info on PSLQ here [14]) (All of the previous fields are from [9]).

Ising integrals, tanh-sinh integration algorithm, box-integrals, minimal polynomials, algebraic numbers in poisson potential functions Needs Citation [4].

7.5.1   BLAME Analysis

This tool is created from LLVM and analyzes code passed through it for operations which can be run at reduced precision. [17]

Complete ?

7.6   Algorithmic concepts dealing with numerical precision

Condition number, the condition number of a function with respect to an argument measures how much the output value of the function can change for a small change in the input argument. Needs Citation

7.7   Theoretical Tools to approach the problem of numerical precision

Here I go over some theoretical tools to confront numerical precision issues. The idea is that some or all of the ideas in this section can be turned on computational code to produce uncertainty estimates of their results.

Complete

7.7.1   Interval arithmetic

This technique focuses on calculating everything with an interval rather than a single number. This way the uncertainty interval is propogated all the way to the end of the calculation. (There is a wiki article but I need to find a real source.) Needs Citation

Unfortunately, it's runtime cost is enormous [14].

Complete

8   Goals for reproducibility

I want to make the following methods more easily executable. Replicability and Compatibility will be huge steps towards making this happen.

Complete

8.1 Method Appropriation

1. **A** is working on airfoil using method $\boldsymbol{\alpha}$, but sim has a problem
2. **B** publishes method $\boldsymbol{\beta}$ which claims to fix said problem
3. **A** reads **B**'s paper
4. **A** downloads **B**'s materials, verifies the result
5. **A** then uses method $\boldsymbol{\beta}$ to solve their problem
6. **A** publishes their airfoil citing **B** as simulation method

8.2 Method Tweaking

1. **A** needs to simulate a large protein to find binding sites
2. **A** finds method $\boldsymbol{\beta}$ by **B** for simulating large proteins efficiently, but doesn't find binding sites.
3. **A** download's **B**'s materials, verifies the result
4. **A** adds code to find binding sites to $\boldsymbol{\beta}$ producing method $\boldsymbol{\alpha}$
5. **A** publishes method $\boldsymbol{\alpha}$ citing **B**'s method $\boldsymbol{\beta}$.

8.3 Method Improvement

1. Performing lattice QCD, **A** finds method $\boldsymbol{\beta}$ by **B**.
2. **A** spots mistake in method $\boldsymbol{\beta}$ from article
3. **A** downloads **B**'s materials, verifies the result
4. **A** changes $\boldsymbol{\beta}$ creating method $\boldsymbol{\alpha}$ solving mistake
5. **A** runs method $\boldsymbol{\alpha}$ verifying better agreement with experimentally measured values
6. **A** publishes $\boldsymbol{\alpha}$ citing **B** about this change

9 Empirical Reproducibility Studies

Here I'm collecting a list of all empirical studies about reproducibility in the scientific community. These are good places to grab statistics to show people how bad the whole situation really is. I'm collecting their timeframe, and what field they cover. See Table 3 for the table.

Complete

Table 3: A list of replicability and reproducibility studies and what they say

Citation	Title	Timeframe	Field	Results
[18]	"Replication in Empirical Economics: The Journal of Money, Credit and Banking Project"	1986	Economics	Response rates to requests for data were higher for articles which were accepted and under review than for articles which were already published (pub: 66%, acc: 96%, rev: 75%), Only 15% of received data sets were properly documented and cited, 26% were incomplete Read More
[19]	"How accurate is scientific software?"	1994	Geophysics	N-version replication study. Approximate cumulative precision-loss of 1% every 4000 source lines (about 3000 executable lines) in terms of average absolute difference.
[20]	"Reproducible Epidemiologic Research"	2005	Epidemiology	93% of articles do not detail data processing. 84% of articles did not release data.
[21]	"The poor availability of psychological research data for reanalysis."	2006	Psychology	After generous attempts to get data from published study authors, only these authors were only able to obtain 25.7% of the datasets in their article cohort.
[22]	"Trial watch: Phase II failures: 2008–2010"	2011	Biology	Success rates in Phase II trails have fallen from 28% to 18%. As mentioned in [23]
[23]	"Believe it or not: how much can we rely on published data on potential drug targets?"	2011	Biology	In only 20-25% of projects were the relevant published data completely in line with their in-house findings
[24]	"Drug development: Raise standards for preclinical cancer research"	2012	Cell Biology	Only 6 of 53 landmark biotechnology papers held up.
[25]	"Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not""	2015	Economics	22/67 (33%) replicated without author contact 29/59 (49%) replicated with author contact.

Table 3 To be continued...

Table 3 (continued)

Citation	Title	Timeframe	Field	Results
[26]	"Estimating the reproducibility of psychological science"	2015	Psychology	97% of original studies had statistically significant results 36% of replications had statistically significant results. 47% of original effect sizes were in the 95% confidence interval of the replication effect size. 83% of studies had a larger effect size in the original study. 39% of effects were subjectively rated to have replicated the original results. Replication success was more consistently related to the original strength of evidence (such as original P value, effect size, and effect tested) than to characteristics of the teams and implementation of the replication (such as expertise, quality, or challenge of conducting the study)
[27]	"Repeatability in Computer Systems Research"	2016	Computer Science	44% of computer science article authors released their source code. Read there are others.
[28]	*Open Data: the researcher perspective - survey and case studies*	2017	Datascience?	This article contains many good statistics. Read Here are a few: 73% of respondants said that having access to others research data benefits or would benefit their own data. Only 65% of respondants said they had shared their research data with others and only 52% said that they provide their research data to publishers so it can be made access with their article. An interesting divide.

10   Reproducibility Software

Here I'm collecting a list of all reproducibility software I have encountered. I'll discuss each in turn hitting what they offer and what they lack. Any specific system which doesn't have it's own section means I haven't tried it or don't know enough about it to say anything.

Complete

10.1   Containers

10.1.1   Introduction

Containers are a method of operating system virtualization that isolate applications including their runtime environment and all the files required to run. LXCs (Linux Containers) are able to run multiple Linux containers on a host with a single kernel. This is done by providing the linux kernel's cgroups feature and isolated namespaces for and isolated environment for containers[29].
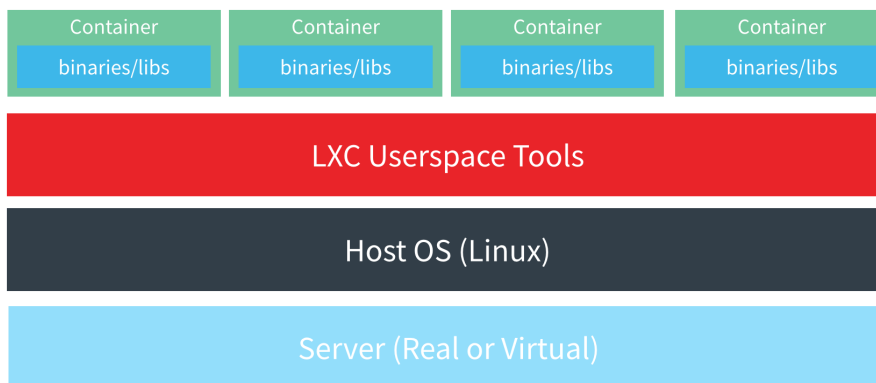
Figure 1: Figure showing a Linux container stack

Containers are restricted to Linux since containers need to be able to isolate processes without a new kernel. They run on the host OS's kernel and rely on system calls for processes. This is stable over different kernel versions and hence the host does not need to install a new kernel depending on the container required version[30].

Once an application is compiled, the set of system calls that an application uses is embedded in the binary. Containers don't abstract the need for the user space and kernel space to share a common set of system calls as seen in figure 2. In a containerized world, this user space is bundled up and shipped around to different hosts, ranging from laptops to production servers.
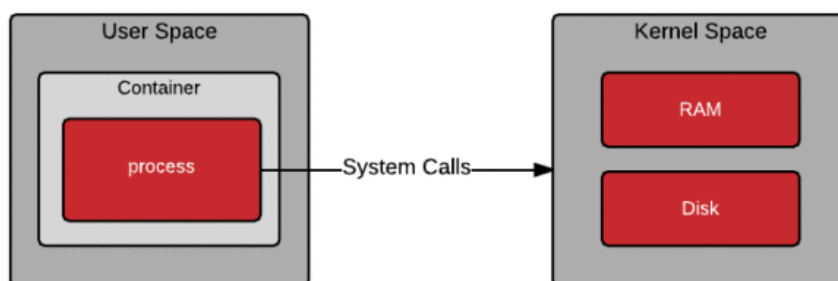
Figure 2: Figure shared system calls

10.1.2   chroot system calls

The chroot system call is a critical foundation for containers. The chroot mechanism can redefine the root directory for any running program, effectively preventing that program from being able to name or access resources outside that root directory tree. These partitions are referred to as chroot jails[31]. It redefines the

root directory for running a program and prevents the program from viewing resources outside the directory. It shares a single kernel with the host while being tamper-resistant and preventing malicious code to "break out" of the jail.

10.1.3 cgroups

cgroups is a Linux kernel feature that allows the user to allocate and monitor resources such as memory, network bandwidth, CPU, disk I/O, etc into hierarchical groups[32]. Some important features provided by cgroups are[33]:

- Resource limiting

- Prioritization

- Accounting

- Control

10.1.4 namespaces

Namespaces wrap a global system resource into an abstraction that makes it appear to the processes within the namespace to appear to have their own isolated instance of the global resource[34].

There are currently 6 types of namespaces[34]

- Mount namespaces - Mount points

- UTS namespaces - Hostname and NIS Domain name

- IPC namespaces - System V IPS, POSIX message queues

- PID namespaces - Process IDs

- Network namespaces - Network devices, ports, etc.

- User namespace - User and group IDs

10.1.5 Security

Initially, in LXCs, the root user of the container could run any code on the host system with complete root access to the host's system. However, since v1.0, multiple security features have been implement to limit access to the host's hardware[35].

10.1.5.1 Discretionary Access Control

It is a method of restricting access to objects based on the identity of subjects and/or groups to which they belong. The controls are discretionary in the sense that a subject with a certain access permission is capable of passing that permission. Based on the filesystem ownership and permissions, the kernel uses discretion to allow access to files[36].

10.1.5.2 Mandatory Access Control

It is a system wide policy that dictates which who is allowed access to certain files and the permissions cannot be altered by the user unless the MAC policy modified or removed.

10.1.6 Popular Containerization Software

- Docker [37]

- Vagrant Needs Citation

- Singularity [38]

- Kuberneties Needs Citation

- Shifter Needs Citation

Complete

10.2 Virtual Machines

These techniques use a hypervisor to run a logically separate virtual machine on a host machine. These tend to be slower, but as hardware vendors add support for virtualization technology to their chipsets they are getting closer to native performance.

- VMWare Needs Citation

- Virtual Box Needs Citation

- KVM Needs Citation

Complete

10.3 Research Environments

From [39].

- Verifiable Computational Research Needs Citation

- knitR Needs Citation

- Collage Authoring Environment Needs Citation

- Sumatra Needs Citation

- Galaxy Needs Citation

- SHARE Needs Citation

- Sweave Needs Citation

- SOLE Needs Citation

- GenePattern Needs Citation

- torch.ch Needs Citation

- Cyverse Needs Citation

- Open Science Framework Needs Citation

- IPOL Needs Citation

- Jupyter Needs Citation

- NanoHUB Needs Citation

- Vistrails Needs Citation

10.4   Workflow Managers

These softwares purport to handle all of our scientific workflow needs. They handle your source code, build your application, run it and compile the paper from the results produced. Some workflows may not fit, but they are working hard to make sure that it doesn't stay that way.

- Popper [40]

- WholeTale Read Categorize Needs Citation

- CodeOcean Read Categorize Needs Citation

- Taverna Needs Citation

- Kurator Needs Citation

- Wings Needs Citation

- Kepler Needs Citation

- Pegasus Needs Citation

- Everware [41]

- CDE Needs Citation

- binder.org Needs Citation

- NextFlow Needs Citation

- Common Workflow Language Needs Citation

- revisit [42]

- workflowr Needs Citation

10.4.1   Popper

Popper [40] is an interesting technology standard which was produced by Ivo Jimenez and his team. It proposes a set of 'requirements' which a piece of research must satisfy in order to be truely reproducible.

The Popper authors correctly point out that source code alone is not enough for complete article replication. Really lots of information is necessary for code to run properly and have a chance of producing the published results.

Popper outlines a set of functionality necessary to allow reproducibility to happen. They provide a simple kind of reference implementation as well which operates off of the git backend. They follow a DevOps philosophy and don't want to re-invent the wheel. Thus using git as a backend they are able to create a fairly robust tool quickly.

Their reference implementation of Popper also includes a number of experiment templates since the learning curve for git and other DevOps tools is quite high. This reduces the barrier to entry a bit.

Complete

10.5   Workflow Analyzers

These software tools glean information about the provenance of computations by analyzing either software meta data or software running on a system.

- FRAPpuccino [43] Read Categorize

- Why-Diff [44] Read Categorize

- YesWorkflow [44] Read Categorize

- NoWorkflow [44] Read Categorize

Complete

10.6   Watchers

These programs use a system like `strace` to track what the researcher is doing and store it in a container. This container an then in-principle be used on any other machine.

- ReproZip [45]

- Prune [46]

- Umbrella [47]   Read   Categorize

- Makeflow [48]   Read   Categorize

- CDE   Needs Citation

- PTU   Needs Citation

- CARE   Needs Citation

  Complete

10.6.1   ReproZip

ReproZip is a kind of 'without-forthought' reproducibility engine which is currently endorsed by the SIGMOD Reproducibility Review. The authors claim that systems in the style of ReproZip but particularly ReproZip promise the following benefits above other types of reproducibility software[45]:

- *portability*: ReproZip claims to eliminate possible portability problems including kernel incompatibility by providing so-called 'unpackers' which can create a VM or container using for example Docker. These packages can then be run on machines with incompatible kernels.[45]

- *extensibility*: ReproZip claims that all which is needed to add support for new platforms is the creation of a new 'unpacker'.[45]

- *modifiability*: ReproZip claims to clearly identify input files so researchers can target the right files to modify input parameters.[45]

- *Usability*: ReproZip allows researchers to customize the created package. This will allow you to remove extra files which you know aren't relevant.[45]

How does ReproZip work? There are two steps. A 'packing' step where necessary files are detected using system call tracing. As this tracing requires `ptrace` currently only linux based OSs are supported[45]. (Is this really that portable then??) [45] contains a bit more information about what commands specifically are sniffed. ReproZip then takes this file information and interfacing with the system package manager figures out which dependencies are required.

With the package created using the process above, ReproZip then invokes an 'unpacker' to create a package which can run on another system. Two supported unpackers use Vagrant and docker. These create vagrant and docker images necessary to run the packaged code from the previous step. These can then be run (ideally) on any machine. They also include a 'directory' and 'chroot' unpacker which work directly out of the filesystem on a machine. The 'directory' unpacker only works when relative filepaths are used, and the 'chroot' version works even when absolute paths are given. (I'm not sure if these other two methods require kernel compatibility or not. It sounds like it does.) [45]

They show an example which creates a docker package of 500MB and a vagrant package of 1.4GB. I actually don't really think these are that small...

The paper also mentions that they are working on distributed code support such as code which uses Hadoop or MPI. [45].

10.6.2 Umbrella

10.6.3 PRUNE

From Doug Thain, this is a different approach which does not monitor system calls to produce reproducible results.

They claim only a 1% overhead.

This method is very similar to a git repository. The user explicitly adds input data or programs and describes their workflow at each 'commit'. Each 'commit' contains 'file's, 'task's, 'result's, and 'environment's.

Here is the paper's description of an Environment: 'An Environment is an explicit statement of the hardware and software needed to execute a Task. An Environment can take many forms, depending on the technology used and the priority placed on reproducibility compared to convenience.' Even a tarball on top of an assumed operating system can be considered an Environment. It appears that this tool uses the 'Umbrella' tool as well.

Environments are created infrequently much like OSs are changed infrequently.

I like the idea of this method, however it seems to require lots of user intervention like git. It would profit from being coupled with an Environment system which I expect to be Umbrella as cited in the paper.

Non-determinism is supported in this scheme by the so-called DBID and CBID ids. Each CBID is created as a hash of the file content while each DBID is available for output files of the process. Thus several files with different CBIDs can be represented by a single DBID.

Apparently PRUNE is managed from a python file which imports the PRUNE library. I'm not sure how I feel about this.

10.6.4 In-Practice

The only one of these schemes which I've tried so far is ReproZip [45]. I ran a simple `htop` command on my laptop and tried to run it on the campus cluster where `htop` doesn't exist. This failed, but I didn't record the error that occured.

Complete

10.7 Advanced Package and Environment Management

These programs take control of the execution environment while also staying out of the way by using existing mechanisms to do their magic. For example, using `LD_LIBRARY_PATH` or `RUNPATH` instead of more complex stuff like userspace namespacing like what containers do.

- Spack [49]

- Nix Needs Citation Expand

- Hashdist Needs Citation Expand

- EasyBuild [50]

10.7.1 Spack

Spack [49] is the only one I've tried so far. It's the best by far.

Complete

10.8 Machine Verification

This software is part of the Software verification ecosystem. This software aims to prove claims about software such as 'This software will never crash'.

- Verified Software Toolchain (VST) (A collection of software) [51]

10.9 Generalized systems

- Reproducible-builds.org [52] Read Expand

- Sciunits [53] Read Expand

Complete

11 Software/Data Repositories

This is a list of the various code and data repositories which currently exist and are trying to solve some part of the replicability problem.

- Computer Physics Communications [54] Read Expand

- DARPA Open Catalog [55] Read Expand

- DOE Code [56] Read Expand

- Mendeley [57] Read Expand

- IPOL [58] Expand

- CodeOcean [59] Read Expand

- researchcompendia.org Needs Citation Read Expand

- ahay.org Needs Citation Read Expand

- osf.io Needs Citation Read Expand

- zenodo.org (Offers integration with some services such as github and JPP(?)) Needs Citation Read Expand

- OCCAM Needs Citation

- DataCenterHub Needs Citation

- RCloud Needs Citation

- RunMyCode.org Needs Citation

- TheDataHub.org Needs Citation

- ChameleonCloud [60]

- Madagascar Needs Citation

- hivebench Needs Citation

- DataSearch Needs Citation

- Data Lighthouse Needs Citation

- RDA (Research Data Alliance) Needs Citation

- Scholix Needs Citation

- Force11 Needs Citation

- o2r [61]

- CALGO [62]

Complete

12 Good examples of reproducible papers

This section is to contain good examples of reproducible papers or good sources of reproducible papers we have encountered.
Complete

- http://stat.columbia.edu/ cunningham/

- http://arokem.org/reproducibility/index.html

- https://trilinos.org ?

13 Interesting Blog Posts

This section is meant to point out some good blog posts

- Adventures in replicable scientific papers: Docker [63].

Complete

14 Legal and Licensing

This section is meant to contain anything related to the legality of reproducing another scientists work.
Complete

14.1 Copyright

Copyright is first defined by the U.S. Constitution by [U.S. Const. art. I, § 8, cl. 8] "To promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries.". This actually creates not only copyright, but patents.

Conventionally, copyright is defined as 'following the pen across the page' and means that you cannot just take words you read on a page and use them yourself.

14.2 Licensing

14.2.1 Copy-Left licenses

14.3 Recommendations

15 Policy Efforts

There are numerous policy efforts on a number of fronts from Journals and Conferences, to Funding agencies. This section is meant to coallate such related issues together.

- *Science* efforts to get researchers to publish code and data. Needs Citation

- *SC'XY* Conference has programs like the student cluster competition. [64], [65]

- The ACM has started offering badges for articles which are replicated or reproduced [66].

Complete

15.1 Reproducibility specific panels or meetings

Here I'm gathering a list of Meetings and panels at which reproducibility is specifically talked about.
Complete

- NSF Workshop on Robustness, Reliability, and Reproducibility in Scientific Research (last meeting 2017)

-

Complete

15.2 Journal Policies

Here I'm putting instances of journal policies which exemplify a commitment to reproducibility, or are good examples of poor policies which will not yield reproducible articles.

In the journal of Pure and Applied Chemistry, guidelines for publication and presentation of results are described circa 1998 in [7], [8].
Complete

16 Community Direction

This section describes people's opinion of the community and where it is going.
Complete

17 Closed Vs Open Source

This section describes my perceived benefits and negatives of closed versus open source development. It is also a spot to discuss any subject relating to Closed or Open Source.
Complete

17.1 Company interest and Lives

The great article [67] discusses the danger of having propietary code running on medical implants. These implants control people's bodily functions or may be relied on to restore normal function in the event of failure. This code can often have bugs inside which can lead to death for the user if serious enough.

Not all users are capable of reading and understanding code in these devices, but those who do have a large interest in making sure the code is as bug-free as possible. Allowing knoweldgeable users to view the code and make bug-fixes would make these devices much safer.
Complete

18 Statistical Tests

This section outlines a few statistical tests which we have used in our analysis.

18.1 Wilcoxon Ranked Sum Test

This test was used in our first JCP paper to test whether the informed and uninformed groups were behaving in a statistically different way. [68], [69]

18.2 Pearson's Chi-squared test for count data

This test was used in our first JCP paper to test the difference between the informed and uninformed groups. [70], [71].

19   Notable Authors

Here I collect together sources from notable authors in the field. I also try to summarize their contributions. Their ideas should be commonly cited in the field, or they should have multiple publications which have done some good work making clear the state of the field.

Complete

19.1   Ioannidis

Complete

19.2   Victoria Stodden

Victoria has been active in the reproducibility community for at least a decade and has contributed a number of interesting pieces of research as well as general recommendations to the community to increase the reproducibility of published articles.

One of her main focuses is in the licensing aspect of reproducibility. She advocates for non-copy-left licenses so that others may freely read and use your published work without worrying about having to change their own code license [72].

She has also been instrumental in sounding the alarm about current scientific practices and how they affect reproducibility of work [73].

Complete

19.3   Claerbout

Complete

19.4   Donoho

Complete

19.5   David H. Bailey

David Bailey concerns himself with numerical precision issues. He has produced a nice summary of scientific problems which require greater than normal precision [4], [9]. He has also been involved in the creation of tools to help find parts of a program which can be run at lower precision ostensibly to limit program execution time[17]. He has also created the mpfun library for thread-safe arbitrary precision arithmetic[16].

He hosts his website[74] which contains all of his papers, talks, and links to code and projects that he works on.

Complete

19.6   Les Hatton

Les deals primarily with software quality and has had a number of studies over the years.

Complete

20   Reference Descriptions

Here I collect together the references I've encountered and read at least some of. I also include a small explanation for each reference. If available, I also break it down by subject.

Complete

20.1   Replicability and Reproducibility Studies

"Reproducible Epidemiologic Research" [20]
American Journal of Epidemiology (2006)
`kwj093.pdf`

This paper describes the types of papers published in 2005 in the 'American Journal of Epidemiology' and the 'Journal of the American Medical Association'. I.e. what kind of study they were, did they use code, and whether that code was available. They find that 93% of articles which required data processing did not discuss this processing. 84% of articles which produced original studies did not report that the data produced was available.   Read

"Is Economics Research Replicable?   Sixty Published Papers from Thirteen Journals Say "Usually Not"" [25]
Finance and Economics Discussion Series 2015-083 (2015)
`2015083pap.pdf`

This paper describes the state of replicability in Economics literature as of 2015. 22/67 (33%) of papers are replicated without contacting the authors, 29/59 (49%) are replicated by contacting the authors.   Read

"Drug development: Raise standards for preclinical cancer research" [24]
Nature (2012)
`483531a.pdf`

An article talking about the need to increase the quality of pre-clinical research to increase the effectiveness of cancer studies and getting working treatments from them. There is focus in the article about how articles which were reproduced were generally of higher quality than those which were not reproduced. They documented more abou their setup and paid closer attention to things like reagents. The authors also outline how pressures from the community in order to publish simple, clear, high-impact results encourages poor practices such as omiting data which doesn't fit their narrative. They also report empirical work by the author and the Amgen company to replicate work from 53 landmark studies. Only 6 of these studies held up.   Read Categorize

"Believe it or not: how much can we rely on published data on potential drug targets?" [23]
Nat Rev Drug Discov (2011)
`printz-rely-published-2011.pdf`

An Empirical study about drug development reproducibility. Apparently companies often perform in-house validation of published findings to help them decide which research to invest in and which not to. The author's company has found that only successful in reproducing about 20-25% of research they attempt. They also state that the prestige of the journal did not ensure reproducibility. "Indeed, our analysis revealed that the reproducibility of published data did not significantly correlate with journal impact factors, the number of publications on the respective target or the number of independent groups that authored the publications." Read Categorize

"Replication in Empirical Economics: The Journal of Money, Credit and Banking Project" [18]
The American Economic Review (1986)
`dewald-replication-econ-1986.pdf`

An empirical study of replication in empirical economics. This study is actually similar to the first JCP

study that we are conducting. They asked authors for data and produce a table of results.   Read Categorize

20.2   Numerical Issues

"High-precision floating-point arithmetic in scientific computation" [9]
Computing in Science Engineering (2005)
`high-prec-arith.pdf`

This paper describes different scientific tasks that require greater than 64-bits of precision. Some fields require hundreds or thousands of digits of precision for their results to be meaningful.

Petascale computing and high-precision arithmetic: Applications and challenges [4]
(2015)
`dhb-zurich-hp.pdf`

This is an invited talk summarizing most of David Bailey's work. This centers around numerical precision and applications which need lots of precision. This presentation also presents a number of instances of reproducibility in physics biomedicine and other fields.

The GNU C Library: Rounding [13]
()

A website describing GNU libc's floating point rounding capabilities.

Accuracy and Stability of Numerical Algorithms [12]
(2002)
`Higham_2002_Accuracy and Stability of Numerical Algorithms.pdf`

This book is a theoretical book discussing the accuracy and stability of numerical algorithms as it's title states. For starters it discusses the idea of machine epsilon.   Read

"How accurate is scientific software?" [19]
IEEE Transactions on Software Engineering (1994)
`hatton-scientific-software-accurate-1994.pdf`

Interesting ongoing N-version scientific software accuracy study in the field of seismic data processing. They claim that 'numerical disagreement grows at around the rate of 1% in average absolute difference per 4000 lines of implemented code, and even worse, the nature of the disagreement is nonrandom.' (Find reference to completed study?) Only read the intro and conclusion.   Read More

20.3   Reproducibility

"Fast Reproducible Floating-Point Summation" [15]
Computer Arithmetic (ARITH), 2013 21st IEEE Symposium on (2013)
`ARITH21_Fast_Sum.pdf`

An algorithm for reproducible parallel summation of floating point numbers with varying accuracy and efficiency.

"Measuring reproducibility in computer systems research" [75]
Unpublished report (2014)
`tr.pdf`

A fun article discussing attempts to get source code and build the code. No attempts to run and reproduce the result were made. The Arizona study technical report Read

"Reproducible Research in Computational Harmonic Analysis" [73]
Computing in Science & Engineering (2009)
`Stodden_Reproducibility_Crisis.pdf`

One of Victoria's articles where she and her collaborators describe the problem of computational reproducibility. They are also plugging their own Matlab reproducibility environments. This is also another mention of Jon Claerbout and really reproducible research. (From [45]: In spite of its importance, achieving reproducibility has proved elusive for computational experiments, and this has led to a credibility crisis.)

"The Problem of Reproducibility" [76]
CHANCE (2012)
`The_Problem_of_Reproducibility.pdf`

Ince describes some of the issues with reproducibility

"Experimental Methodologies for Large-Scale Systems: a Survey" [77]
Parallel Processing Letters (2009)
`RR-6859.pdf`

The authors describe some different methodologies in running computational 'experiments'. This includes the classification of hardware and software environments.

"Quantifying Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome" [78]
PLoS ONE (2013)
`journal.pone.0080278.pdf`

Establishes the idea of the reproducibility score which quantifies the cost of reproducing the result in the event that details are left out of publications. Gives a nice in depth review of making a computational biology paper reproducible. Much like what I plan with the Vertical study

"A collaborative approach to computational reproducibility" [79]
Information Systems (2016)
`1-s2.0-S0306437916300813-main.pdf`

A nice short paper describing a system where people create 'reproducibility papers' and report the existence of tools for making environments and libraries and what not available for researchers to use.

"Tools and techniques for computational reproducibility" [80]
GigaScience (2016)
`ToolsAndTechniquesComputationalReproducibilityPicollo2016.pdf`

This paper details several techniques that can be helpful to automate the reproduction of results including things like scripts and using docker containers to create virtualized environments.

"Repeatability of published microarray gene expression analyses" [81]
Nat Genet (2009)
`ng.295.pdf`

    The researchers attempt to reproduce a bunch of microarray gene expression papers. These reproductions were limited by the availability of data and in some cases code.

"Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates" [82]
Proceedings of the National Academy of Sciences (2016)
`PNAS-2016-Eklund-7900-5.pdf`

    The authors discuss how fMRI systems have huge false positive rates which is related to the existence of the dead salmon poster: Bennett-Salmon-2009.pdf

"Where next for the reproducibility agenda in computational biology?" [83]
BMC Systems Biology (2016)
`WhereNextForComputationalBiology.pdf`

    This paper puts up a set of recommendataions for computational biology to facilitate reproducibility. Nothing that new.

"A First Look at Reproducibility and Non-Determinism in CMS Software and ROOT Data" [84]
CMS Technical Report TR-2016-01 (2016)
`repro-tr-2016.pdf`

    An interesting paper which discusses efforts to make a CMS ROOT workflow reproducible.

"Software simplified" [85]
Nature (2017)
`546173a.pdf`

    A paper discussing Containers and how they are taking over the world

"The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements" [86]
PLOS ONE (2012)
`journal.pone.0038234.pdf`

    A great paper talking about scientific output differences between operating systems

"Estimating the reproducibility of psychological science" [26]
Science (2015)
`aac4716.full.pdf`

    An excellent and important article presenting an empirical study about the reproducibility of psychological science. This is a great article for a review of current thinking in the reproducibility world. See section 9 for details about the study in a greater context. More Cites

"Keeping the faith" [87]
Science (2017)

```
service-dinos-keeping-faith-2017.pdf
```

An article describing a possible revolution in paleontology. A paleontologist Mary Schweitzer is using new techniques to dissolve dinosaur bone and retrieve partially intact organic molecules from within. These molecules are then compared to living animals to find information about them.

These are big results as they go against industry understanding of how long organic materials can survive. The primary issue for her however, is that other labs have not been able to reproduce her results. Until this happens she will have a hard time pushing her narrative.

20.4   Benchmarking

"Scientific Benchmarking of Parallel Computing Systems: Twelve Ways to Tell the Masses when Reporting Performance Results" [88]
Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (2015)
```
a73-hoefler.pdf
```

This paper discusses a number of problems with current benchmarking practices in the HPC community. They present their own library for HPC benchmarking as a solution to this problem.

"A high-performance, portable implementation of the MPI message passing interface standard" [89]
Parallel Computing (1996)
```
1-s2.0-0167819196000245-main.pdf
```

One of the founding papers of MPI, showing how MPI has been able to create a message passing system which can achieve high performance even compared to vendor specific implementations on a variety of systems.

"I/O in Parallel Applications: the Weakest Link" [90]
International Journal of High Performance Computing Applications (1998)
```
International Journal of High Performance Computing Applications-1998-Thakur....
```

Goes into the problem of data throughput in HPC.

"Test suite for evaluating performance of multithreaded {MPI} communication" [91]
Parallel Computing (2009)
```
1-s2.0-S0167819109000143-main.pdf
```

This is a more detailed article about MPI benchmarking. Specifically, this paper deals with threading performance, and the challenges therein.

"Reproducible Measurements of MPI Performance Characteristics" [92]
Proceedings of the 6th European PVM/MPI Users' Group Meeting on Recent Advances in Parallel Virtual Machine and Message Passing Interface (1999)
```
mpptest.pdf
```

Describes a number of problems associated with testing parallel message passing implementations. Describes their test suite and how they get around these problems.

20.5 Legal and Licensing

"The Legal Framework for Reproducible Scientific Research: Licensing and Copyright" [72]
Computing in Science Engineering (2009)
`stodden-legal-framework-reproducible-2009.pdf`

A nice article by Victoria discussing copyright and the various licenses which deal with releasing copyright and how it applies to reproducible science. The article makes the case that copy-left licenses prevent the sharing of code since people often have worries or concerns about licensing, or want to produce products from what they learn from published works. If code includes copy-left provisions it will have to be passed up.

She concludes with the recommendation that non-share-alike licenses are necessary for all types of work from code to data releases.

"Pay up or retract? Drug survey spurs conflict" [5]
Science (2017)
`marcus-payup-retract-2017.pdf`

A great article describing a scandal currently underway being perpetrated by Donald Morisky of the University of California, Los Angeles. At some point he produced the Eight-Item Morisky Medication Adherence Scale (MMAS-8) which is measured from a questionaire administered to patients taking a drug. This scale is supposed to measure how well patients are following the doctor's directions. This scale is often used as part of drug testing studies in order to see the effect of patient faithfulness to a drugs effectiveness.

The problem is that he asks for large licensing fees to be paid in order to use this questionaire. Fees are often on the order of $1200-$1500. This article details the plight of a specific graduate student who after asking Morisky for permission to use the scale by email did not hear back from him. She went ahead anyway, and now they're asking for $6500 for use of the scale otherwise she will have to retract her paper.

20.6 Policy Efforts

"Incentivizing Reproducibility" [66]
Commun. ACM (2016)
`p5-boisvert.pdf`

Defines the 'Results Replicated' and the 'Results Reproduced' badges for the ACM. Read More

 [64]
ACM SIGHPC Connect (2016)
`CONNECT_14_Oct2016.pdf`

Newsletter describing the student cluster competition at SC16. Read More

 [65]
ACM SIGHPC Connect (2017)
`CONNECT_15_Feb2017.pdf`

Newsletter describing the winner of the reproducibility initiative at SC16 and the paper for the student cluster competition for SC17. Read More

"Replication, Replication" [93]
PS: Political Science and Politics (1995)

```
replication.pdf
```

A nice article V.S. Recommended which details various issues faced by researchers in empirical political sciences to replicate works by other authors. They include the rather surprising statistc from another paper that 90.1% of political science articles are never cited! The article ends in a discussion of policy changes to make to improve reproducibility in political science. I did not finish reading the article. More Cites Read More

20.7   Reproducibility Benchmarking

"Reproducibility in density functional theory calculations of solids" [94]
Science (2016)
```
aad3000.full.pdf
```

A very detailed study in which a set of DFT codes had their reproducibility benchmarked. The developers and researchers for these codes got together and decided on a figure of merit to quantify their reproducibility.

20.8   Uncertainty Quantification

"Validation and Uncertainty Assessment of Extreme-Scale HPC Simulation through Bayesian Inference" [95]
Euro-Par 2013 Parallel Processing: 19th International Conference, Aachen, Germany, August 26-30, 2013. Proceedings (2013)
```
chp%3A10.1007%2F978-3-642-40047-6_7.pdf
```

20.9   Sources of Variation

"A Critical Assessment of Storytelling: Gene Ontology Categories and the Importance of Validating Genomic Scans" [96]
Molecular Biology and Evolution (2012)
```
mss136.pdf
```

A discusion of analysis methods in biology used to create a narrative for specific genes and whatnot. That non-reproducibility can so easily affect the result of analysis is the main point of this paper. Not about the reproducibility crisis in biology...

"Offspring from Reproduction Problems: What Replication Failure Teaches Us" [97]
()
```
P13-1166.pdf
```

This paper dicusses attempts to replicate results from word similarity programs. They discuss several sources which can be sources of error

"Design and Analysis of Computer Experiments" [98]
Statist. Sci. (1989)
```
euclid.ss.1177012413.pdf
```

An old paper describing a system for studying variation of computational simulations given different initial conditions and assuming deterministic simulations. This is a good starting point for my vertical research!

"Measuring and Understanding Extreme-Scale Application Resilience: A Field Study of 5,000,000 HPC Application Runs" [99]
2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (2015)
`07266835.pdf`

This paper describes the DEPEND team's tool LogDiver and their work studying errors which programs encounter on large supercomputers.

Resolving Numerical Anomalies in Scientific Computation [14]
(2008)
`dhb-numerical-bugs.pdf`

A nice article describing different numberical precision issues which programs may encounter.

20.10   Software Quality

"Are all global alignment algorithms and implementations correct?" [100]
bioRxiv (2015)
`031500.full.pdf`

This paper discusses the existence of bugs in implementations of global alignment algorithms

"The State of Software in Evolutionary Biology" [101]
bioRxiv (2015)
`031930.full.pdf`

This paper discusses the code quality of several bioinformatics tools which are used in academia They find some problems and propose some practices for the scientific community.

"High-Integrity Software, Computation and the Scientific Method" [102]
(2011)
`Hatton_ICALEPCS2011.pdf`

This article deals with several issues which can undermine a scientific result from software. It takes fault rates specifically which result in subtly incorrect results from scientific programs.

"The Pitfalls of Verifying Floating-point Computations" [103]
ACM Trans. Program. Lang. Syst. (2008)
`a12-monniaux.pdf`

A deep case based analysis of floating point reproducibility problems which can arrise in the analysis of computational code. (a lengthy read)

Code Complete [104]
(2004)
`code-complete-2nd-edition-v413hav.pdf`

A large book talking about good coding practices. Recommended by Victoria's Friend

Software Engineering – Product Quality [105]
(2011)
```
A
```

n ISO standard describing aspects of software quality so that the quality of disparate types of software can be compared on equal footing.

"The Chimera of Software Quality" [106]
Computer (2007)
```
hatton-software-quality-2007.pdf
```

A general report about software quality from Hatton. He dicusses fields he has worked on and the prevalence of computing errors and people's refusal to accept that they produce buggy code.

20.11   White Paper References

MPI: A Message-Passing Interface Standard [107]
(1994)

The article defining 1.0 of the MPI standard.

"OpenMP: An Industry-Standard API for Shared-Memory Programming" [108]
IEEE Comput. Sci. Eng. (1998)

This paper describes the OpenMP Standard

"ENZO: An Adaptive Mesh Refinement Code for Astrophysics" [109]
ApJS *(2014)*

This paper describes Enzo and is the official Enzo citation.

"The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores" [110]
The Astrophysical Journal (2013)
```
Behroozi_2013_ApJ_762_109.pdf
```

The rockstar halo finding algorithm.

"The Spack Package Manager: Bringing Order to HPC Software Chaos" [49]
Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (2015)
```
a40-gamblin.pdf
```

The official citation and introduction to Spack

Hierarchical data format version 5 [111]
(2000-2010)

The official citation of HDF5

"yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data" [112]
ApJS *(2011)*

The official method paper and citation of YT

"ReproZip: Computational Reproducibility With Ease" [45]
Proceedings of the 2016 International Conference on Management of Data (2016)
`p2085-chirigati.pdf`

This is the official paper describing the ReproZip tool. This paper by all means has a GREAT introdution to the various types of software to enforce reproducibility. Please see 10.6.1 for an in-depth discussion of this tool and paper.

reproducible-builds.org [52]
()

This is the reproducible-builds.org website. They give information about how to make your software reproducible.

"Sciunits: Reusable Research Objects" [53]
(2017)
`1707.05731.pdf`

I need to read this paper.

Computer Physics Communications [54]
()

A reference with a link to the Computer Physics Communications Journal

DARPA Open Catalog [55]
()

A reference with a link to the DARPA Open Catalog

DOE Code [56]
()

A reference with a link to the DOE code Catalog

Mendeley data repository for Comptuter Physics Communications [57]
()

A reference with a link to the mendeley Computer Physics Communications repository

"Floating-point Precision Tuning Using Blame Analysis" [17]
Proceedings of the 38th International Conference on Software Engineering (2016)
`blame-analysis.pdf`

This paper describes the BLAME Analysis tool which examines a program's instructions and determines which operations can be floating point and which can be double. It runs once and is faster than other tools such as 'Precimonious' which was mentioned.   Read More

A Thread-Safe Arbitrary Precision Computation Package [16]
()
`mpfun2015.pdf`

A paper describing the mpfun tool which is used to provide thread-safe arbitrary precision arithmetic.

"Docker: Lightweight Linux Containers for Consistent Development and Deployment" [37]
Linux J. (2014)

A web article describing Docker and its use for consistent development and deployment. Should use as a white paper   Read

"Singularity: Scientific containers for mobility of compute" [38]
PLOS ONE (2017)
`Singularity.pdf`

A White paper describing singularity and it's functionality.   Read

"Jupyter Notebooks – a publishing format for reproducible computational workflows" [113]
Proceedings of the 20th International Conference on Electronic Publishing (2016)
`JupyterNotebooks.pdf`

A white paper describing Jupyter notebooks and its capabilities as a reproducible research platform. Read

"PRUNE: A preserving run environment for reproducible scientific computing" [46]
2016 IEEE 12th International Conference on e-Science (e-Science) (2016)
`Thain-Prune-2016.pdf`

A Thain article about PRUNE a tool for preserving run environments for reproducible scientific computing. This is a white paper describing their tool. It is a kind of workflow management system which tries to combine aspects of workflow managers such as NextFlow with version control systems such as git.

R: A Language and Environment for Statistical Computing [114]
(2008)

The official citation for the R language.   Read

"Standing on the Shoulders of Giants by Managing Scientific Experiments Like Software" [40]
;login: The USENIX Magazine ()
`jimenez-login16.pdf`

This is a kind of white paper which describes the Popper reproducibility methodology. A couple items of interest: The authors note correctly that source code itself is often not enough for a successful replication of their work. There are issues like necessary library or language versions which will prevent exact duplication of their results.

Popper is designed to be a standard of some sort which can be implemented with currently available tools. Any set of tools which satisfies the outlined functionality can be used. In fact in their paper, they detail at least 2-3 existing technologies which can satisfy the requirements of specific steps in the reproducibility process.

"revisit: a Workflow Tool for Data Science" [42]
arXiv (2017)
`matloff-revisit-2017.pdf`

White paper describing the revisit workflow system Read Categorize

20.12   Unassigned

"A Critical Review on the Use of Support Values in Tree Viewers and Bioinformatics Toolkits" [115]
bioRxiv (2017)
`035360.full.pdf`

This paper discusses some of the problems with philogenetic tree programs. they had a hard time reproducing published trees.   Categorize

"Full Computational Reproducibility in Biological Science: Methods, Software and a Case Study in Protein Biology" [116]
(2016)
`1608.06897.pdf`

An interesting paper by Les Hatton talking about full reproducibilty in biology.   Read Categorize

"The AGORA High-resolution Galaxy Simulations Comparison Project. II. Isolated Disk Test" [117]
The Astrophysical Journal (2016)
`1610.03066.pdf`

An extensive test of reproducibility of astrophysical codes across code types.   Categorize

"Enhancing reproducibility for computational methods" [118]
Science (2016)
`1240.full.pdf`

Read Categorize Complete

"What Computational Scientists Need to Know About Intellectual Property Law: A Primer" [119]
Implementing Reproducible Research (2014)
`Stodden_chapter.pdf`

   What computational scientists need to know about intellectual property law: A primer Read Categorize

"Making massive computational experiments painless" [120]
2016 IEEE International Conference on Big Data (Big Data) (2016)
`osbg-MDS2016.pdf`

   Clusterjob provides an interface to break work down into jobs which can be farmed out to a multitude of interfaces. Read Categorize Complete

"Contradicted and initially stronger effects in highly cited clinical research" [121]
JAMA (2005)
`JOC50060.pdf`

   A great paper describing replication outcomes for clinical trials Categorize

"Repeatability in Computer Systems Research" [27]
Commun. ACM (2016)
`p62-collberg.pdf`

   A nice paper about Repeatability in Computer systems research Read Categorize

"Why Most Published Research Findings Are False" [122]
PLOS Medicine (2005)
`journal.pmed.0020124.PDF`

   A great article explaining why most published research findings are false. This article discusses some of the career pressures affecting scientist's decisions. It is also a theoretical argument about how more than half of existing psychological research is false. [26]. Categorize

"1,500 scientists lift the lid on reproducibility" [123]
Nature (2016)
`533452a.pdf`

   An article discussing survey responses from scientists asking about reproducibility in their fields. Read Categorize

"Reproducibility in Science" [124]
Circulation Research (2014)
`116.full.pdf`

   An article discussing the root causes of the reproducibility problem. Read Categorize

How Computational Science is Changing the Scientific Method [125]
 (2009)
`Science20July2009VictoriaStodden.pdf`

A talk by V.S. about how computation is affecting the scientific method.   Categorize

Artifact Review and Badging [1]
 ()

A website describing the ACM badging policy. Provides the definitions for Reproducibility terms Categorize

"WaveLab and Reproducible Research" [126]
Wavelets and Statistics (1995)
```
10.1007%2F978-1-4612-2544-7.pdf
```

A paper which originates the term 'really reproducible research'. It's also about the wavelab matlab software collection. Read Categorize

REPRODUCIBLE COMPUTATIONAL RESEARCH: A history of hurdles, mostly overcome [127]
 ()

A website by Claerbout himself discussion his history with reproducible research. Read Categorize

Neuromorphic Chips Are Destined for Deep Learning—or Obscurity [2]
 ()

A website describing neuromorphic chips and how they have applications for deep learning.   Read Categorize

"Electronic documents give reproducible research a new meaning" [128]
SEG Technical Program Expanded Abstracts (1992)

The original wavelab article about reproducible research. Lays the groundwork for Claerbout's wavelab plan. Read Categorize

"Reproducible Research" [129]
Computing in Science & Engineering (2009)

Another clarbout article about reproducibility. Read Categorize

"Three Dream Applications of Verifiable Computational Results" [130]
Computing in Science Engineering (2012)
```
ThreeDreams.pdf
```

An Article by Donoho describing three dream applications of verifiable computational results. Similar to my Three Methods. Read Categorize

"Exploring Translation of OpenMP to OpenACC 2.5: Lessons Learned" [131]

2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW) (2017)
`TranslationOpenMPOpenACC.pdf`

An article describing how to translate between OpenMP and OpenACC. What is missing, and lessons learned. Not exactly portability.. More like Translation. Read Categorize

"Next-Generation Implications of Open Access" [132]
CTWatch Quarterly (2007)
`ProgressIn10Years.pdf`

This article contains predictions for where technology and the community will be in 10 years time. Interestingly, it seems like fusion. We're still 10 years away from their predictions. Read Categorize

"Facilitating reproducible research by investigating computational metadata" [44]
2016 IEEE International Conference on Big Data (Big Data) (2016)
`computational-meta-data.pdf`

This describes several workflow and metadata analysis tools for work provenance. Why-Diff, YesWorkflow, NoWorkflow Read Categorize

"FRAPpuccino: Fault-detection through Runtime Analysis of Provenance" [43]
9th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 17) (2017)
`hotcloud17-paper-han.pdf`

A paper describing the meta-workflow analysis software FRAPpuccino. Read Categorize

doi:10.1080/01621459.1972.10481279 doi:10.1080/01621459.1972.10481279 doi:10.1080/01621459.1972.10481279 *(***doi:10.1080/01621459.1972.10481279***)*

An old paper describing the Wilcoxon statistical test among others. The paper isn't online, you have to order it. Categorize

10.2307/2984263 10.2307/2984263 10.2307/2984263 *(***10.2307/2984263***)*

An old paper describing Pearson's Chi-squared test for count data. The paper is online, but you have to pay for it. Categorize

Wilcoxon Rank Sum and Signed Rank Tests [69]
()

Implementation of the wilcoxon test in R. Categorize

"Constructing Confidence Sets Using Rank Statistics" [68]
Journal of the American Statistical Association (1972)

Theoretical description of Wilcoxon test. Read Categorize

Pearson's Chi-squared Test for Count Data [71]
 ()

Implementation of the pearson chisquared test in R.   Categorize

"A Simplified Monte Carlo Significance Test Procedure" [70]
Journal of the Royal Statistical Society. Series B (Methodological) (1968)

Contains the theoretical description of the pearson chisquare test.   Read   Categorize

"The Scientific Method in Practice: Reproducibility in the Computational Sciences" [133]
MIT Sloan Research Paper (2010)
SSRN-id1550193.pdf

A discussion from Victoria about reproducibility and the computational sciences.   Read   Categorize

"SC16 Explores Reproducibility for Advanced Computing Through Student Cluster Competition" [134]
 ()

A reference Victoria insisted on having in the SC17 poster.   Read   Categorize

"Facilitating Reproducibility in Scientific Computing: Principles and Practice" [135]
Reproducibility (2016)
vcs-facilitating-reproducibility

A Chapter in a book VCS wrote with Bailey.   Read   Categorize

Workshops to Gauge the Impact of Requirements for Public Access to Data Produced by NSF-funded Research in Mathematics and the Physical Sciences [136]
 (2016)
MPS_Workshop_Report_FinalVersion.pdf

A Report from a group of people who studied the effectiveness of public science policies.   Read   Categorize

"IEEE Standard for Floating-Point Arithmetic - Redline" [11]
IEEE Std 754-2008 (Revision of IEEE Std 754-1985) - Redline (2008)
ieee-754-2008-redline.pdf

The IEEE 754 2008 standard with changes from the 1984 standard shown.   Read   Categorize

"IEEE Standard for Floating-Point Arithmetic" [10]
IEEE Std 754-2008 (2008)
ieee-754-2008.pdf

The IEEE 754 2008 standard.   Read   Categorize

PERFORMANCE DEVELOPMENT [137]
 ()

The top500 website with interesting statistics about supercomputers

"No evidence for or against gravitational waves" [138]
Nature (2014)
`bicep2-no-grav-waves`

A review article in Nature about the beleagered BICEP2 gravitational wave measurement.   Read
Categorize

"A joint analysis of Planck and BICEP2 B modes including dust polarization uncertainty"
[139]
Journal of Cosmology and Astroparticle Physics (2014)

An article describing the lack of strong evidence for BICEP2 gravitational wave measurement.   Read
Categorize

"Toward an understanding of foreground emission in the BICEP2 region" [140]
Journal of Cosmology and Astroparticle Physics (2014)

An article describing the lack of strong evidence for BICEP2 gravitational wave measurement.   Read
Categorize

"High-precision computation: Mathematical physics and dynamics" [141]
Applied Mathematics and Computation (2012)
`dhb-hp-comp.pdf`

   Read Categorize

"High-Precision Arithmetic in Mathematical Physics" [142]
Mathematics (2015)
`dhb-hp-math-physics.pdf`

   Read Categorize

"Precimonious: Tuning assistant for floating-point precision" [143]
*2013 SC - International Conference for High Performance Computing, Networking, Storage and Analysis
(SC) (2013)*
`precimonious.pdf`

   Read Categorize

"Setting the Default to Reproducible" [144]
ICERM Workshop 2013 Report ()
`default-reproducible.pdf`

Read Categorize

"Code Ocean Is Live: Upload Your Algorithms" [59]
IEEE Design Test (2017)
`code_ocean_feb_2017.pdf`

A document announcing the start of codeocean.   Read Categorize

IPOL Journal - Image Processing On Line [58]
()

This is the website of the IPOL journal.   Categorize

"Killed by Code: Software Transparency in Implantable Medical Devices" [67]
Software Freedom Law Center ()
`killed-by-code.pdf`

This is a great article discussing proprietary code running on implantable medical devices.   Read
Categorize

Scandal Rocks Scientific Community [145]
()

A website describing the Schon scandal and some of its fallout.   Categorize

"High-precision computation: Mathematical physics and dynamics" [146]
Applied Mathematics and Computation (2012)
`dhb-hpmpd.pdf`

Read Categorize

"Topological susceptibility with the asqtad action" [147]
Phys. Rev. D (2010)
`PhysRevD.81.114501.pdf`

Read Categorize

"Sparse MRI: The application of compressed sensing for rapid MR imaging" [148]
Magn Reson Med (2007)
`donoho-mri.pdf`

Read Categorize

"Editorial: ACM TOMS Replicated Computational Results Initiative" [149]
ACM Trans. Math. Softw. (2015)
`heroux-replicated-results-initiative.pdf`

Read Categorize

"Creating Repeatable Computer Science and Networking Experiments on Shared, Public Testbeds" [150]
SIGOPS Oper. Syst. Rev. (2015)
`p90-edwards.pdf`

Read Categorize

Adventures in replicable scientific papers: Docker [63]
(2015)

A blog post about docker Read Categorize

OCCAM [151]
()

The website of OCCAM. Read Categorize

"Toward Resilient Algorithms and Applications" [152]
(2014)
`1402.3809.pdf`

Read Categorize

"A Quantitative Study on the Re-executability of Publicly Shared Scientific Workflows" [153]
2015 IEEE 11th International Conference on e-Science (2015)
`mayer-reexecutability-2015.pdf`

Something to do with scientific repeatability Read Categorize

"The hard road to reproducibility" [154]
Science (2016)
`the-hard-road-to-reproducibility.pdf`

Article having to do with challenges of reproducibility Read Categorize

"Has open data arrived at the British Medical Journal (BMJ)? An observational study" [155]
BMJ Open (2016)
`e011784.full.pdf`

Article about the open data policy at BMJ. Read Categorize

TOP Guidelines [156]
()

Open science guidelines from Center for Open Science. Read Categorize

osf-transparency-top-guidelines osf-transparency-top-guidelines osf-transparency-top-guidelines *(osf-transparency-top-guidelines)*

Open Science Foundation TOP guidelines Read Categorize

"A new ensemble-based consistency test for the Community Earth System Model (pyCECT v1.0)" [157]
Geoscientific Model Development (2015)
`gmd-8-2829-2015.pdf`

CESM consistency test. People in the community looking into their application's reproducibility. Read Categorize

"Evaluating statistical consistency in the ocean model component of the Community Earth System Model (pyCECT v2.0)" [158]
Geoscientific Model Development (2016)
`gmd-9-2391-2016.pdf`

CESM consistency test, ocean component. People in the community looking into their application's reproducibility. Read Categorize

"Towards Characterizing the Variability of Statistically Consistent Community Earth System Model Simulations" [159]
Procedia Computer Science (2016)
`CESM-static-consist-2016.pdf`

CESM statistical consistency test. People in the community looking into their application's reproducibility Read Categorize

Dear Colleague Letter: Reproducibility and Robustness of Results [160]
(2016)

A Dear Colleagues Letter about reproducibility and robustness in Geosciences Directorate Read Categorize

Dear Colleague Letter: Robust and Reliable Research in the Social, Behavioral, and Economic Sciences [161]
(2016)

A Dear Colleagues Letter about reproducibility and robustness in Social Behavioural and Economic sciences Read Categorize

Dear Colleague Letter: Encouraging Reproducibility in Computing and Communications Research [162]
(2016)

A Dear Colleagues Letter about reproducibility in Computer and Information Science and Engineering. Read Categorize

A First Look at Reproducibility and Non-Determinism in CMS Software and ROOT Data [163]
 (2016)
`repro-chep-2016.pptx`

   A talk by Thain about his CMS ROOT work   Read   Categorize

"Conducting reproducible research with Umbrella: Tracking, creating, and preserving execution environments" [47]
2016 IEEE 12th International Conference on e-Science (e-Science) (2016)
`Thain-Umbrella-2016.pdf`

   A Thain article about Umbrella tool for creating and preserving execution environments. (PACKAGING)
Read   Categorize

"Verifying Computations Without Reexecuting Them" [164]
Commun. ACM (2015)
`p74-walfish.pdf`

   Article about trying to verify computations without reexecuting them   Read   Categorize

"Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight" [165]
Proceedings of the National Academy of Sciences (2016)
`PNAS-2016-Shneiderman-13538-40.pdf`

   Article about faulty algorithms and independent oversight.   Read   Categorize

"Computing Krippendorff's Alpha-Reliability" [166]
 (2011)
`KrippendorffAlpha-2011.pdf`

   Article by Krippendorff about Krippendorff's Alpha   Read   Categorize

Neurofinder [167]
 ()

   A neuroscience ML benchmark framework. Related to method comparability!   Read   Categorize

"Big Data in Practice" [168]
Digital Journalism (2016)
`Big Data in Practice.pdf`

   Read   Categorize

The World Factbook [169]
 ()

   The CIA Factbook containing among other things, the list of country codes for each country.   Categorize

"Design and Analysis of Computer Experiments" [98]
Statist. Sci. (1989)
`euclid.ss.1177012413.pdf`

Article talking about designing computational experiments, relevant for vertical project Read Categorize

Cyclomatic complexity [170]
()

A wiki article describing cyclomatic complexity. Read Categorize

"Power-Law Distributions of Component Size in General Software Systems" [171]
IEEE Transactions on Software Engineering (2009)
`hatton-software-size-power-law-2009.pdf`

An interesting report about software component size and its power law distribution. Read Categorize

"Reexamining the fault density component size connection" [172]
IEEE Software (1997)
`hatton-fault-component-size-1997.pdf`

An interesting report about how software fault rate depends on software component size. Read Categorize

"The case for open computer programs" [173]
Nature (2012)
`nature10836.pdf`

A case for open computing from hatton and ince. Read Categorize

"Everware toolkit. Supporting reproducible science and challenge-driven education." [41]
(2017)
`everware.pdf`

White paper for Everware. An open science platform? Read Categorize

"MPI Benchmarking Revisited: Experimental Design and Reproducibility" [174]
(2015)
`mpi-benchmarking-revisited-2016.pdf`

Paper on benchmarking MPI software. Focusing on practices for MPI software experiments. Read Categorize

"Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science" [175]
(2015)
`SBE_Robust_and_Reliable_Research_Report.pdf`

NSF report on Robust and reliable research. Read Categorize

"Reproducibility of computational workflows is automated using continuous analysis" [176]
Nat Biotech (2017)

Article about using continuous analysis to make workflows reproducible. Need to get a copy of this article. Read Categorize

Data Sharing and Reproducibility [39]
(2017)
`NASSackler2017-2-STODDEN.pdf`

A Talk with a list of a bunch of tools. Read Categorize

"Portable, high-performance containers for HPC" [177]
(2017)
`shifter-2017.pdf`

A paper describing shifter which is essentially containers for HPC Read Categorize

"Reproducibility in Human Neuroimaging Research: A Practical Example from the Analysis of Diffusion MRI" [178]
()
`the-practice-of-reproducibility.pdf`

A book giving examples of reproducible workflows in practice. Read Categorize

"Reproducible MPI Micro-Benchmarking Isn'T As Easy As You Think" [179]
Proceedings of the 21st European MPI Users' Group Meeting (2014)
`p69-hunold.pdf`

An article detailing MPI benchmarking problems. Read Categorize

"On the Expected and Observed Communication Performance with MPI Derived Datatypes" [180]
Proceedings of the 23rd European MPI Users' Group Meeting (2016)
`p108-carpen-amarie.pdf`

An article describing expected peformance variations with MPI. Read Categorize

"Automatic Verification of Self-consistent MPI Performance Guidelines" [181]
Proceedings of the 22Nd International Conference on Euro-Par 2016: Parallel Processing - Volume 9833 (2016)

A whitepaper about the pgmpi tool for verification of MPI performance. Need to get a pdf of this source. Read Categorize

ReproMPI Benchmark [182]
()

The github repo for ReproMPI. Needs Citation Read Categorize

"Stepping Stones to Reproducible Research: A Study of Current Practices in Parallel Computing" [183]
Euro-Par 2014: Parallel Processing Workshops: Euro-Par 2014 International Workshops, Porto, Portugal, August 25-26, 2014, Revised Selected Papers, Part I (2014)
`europar-2014-part1.pdf`

An article about current practices in reproducible parallel research. Read Categorize

"Research: Publication bias and the canonization of false facts" [184]
eLife (2016)
`nissen-false-facts-2016.pdf`

Interesting article about how false facts can be canonized by reviewer bias. Read Categorize

"Retraction" [185]
Science (2006)
`chang-retraction-2006.pdf`

A retraction by a prominent biophysicist because of an error in an in-house data processing program. Evidence that professional 'eyeball' inspection is not enough and the code must be made available. Categorize

"A Scientist's Nightmare: Software Problem Leads to Five Retractions" [186]
Science (2006)
`miller-scientists-nightmare-2006.pdf`

A discussion of the Chang retraction. Read Categorize

"In-Datacenter Performance Analysis of a Tensor Processing Unit" [3]
Proceedings of the 44th Annual International Symposium on Computer Architecture (2017)
`p1-Jouppi.pdf`

An interesting paper about a new type of silicon, the TPU or tensor processign unit. This is designed for machine learning tasks and is optimized to work with TensorFlow. It is a Google internal device. Read Categorize

"Provisioning Reproducible Computational Science" [187]
(2014)
`xsede14_stodden.pdf`

Read Categorize

"Toward standard practices for sharing computer code and programs in neuroscience" [188]
Nat Neurosci (2017)
`nn.4550.pdf`

A Nature commentary on requiring code to be shared in neuroscience. Read Categorize

"Extending transparency to code" [189]

Nat Neurosci (2017)
`nn.4579.pdf`

A Nature editorial describing their new policy for code transparency in neuroscience.   Read Categorize

"Case studies in reproducibility" [190]
Briefings in Bioinformatics (2011)
`hothorn-reproducibility-case-studies-2011.pdf`

Another article arguing the benefits of looking at data and software for published studies.   Read Categorize

"TriCheck: Memory Model Verification at the Trisection of Software, Hardware, and ISA" [191]
Proceedings of the Twenty-Second International Conference on Architectural Support for Programming Languages and Operating Systems (2017)
`tricheck.pdf`

A whitepaper describing the tricheck tool. A kind of memory checking program (It does more than that though..) Read Categorize

"How to Make a Multiprocessor Computer That Correctly Executes Multiprocess Programs" [192]
IEEE Transactions on Computers (1979)
`lamport-multiprocess-checking-1979.pdf`

A foundational paper describing how to check that a multiprocess program is being executed correctly. Read Categorize

"Next Generation Clouds, the Chameleon Cloud Testbed, and Software Defined Networking (SDN)" [60]
Proceedings of the 2015 International Conference on Cloud Computing Research and Innovation (ICCCRI) (2015)
`chameleon-cloud.pdf`

A whitepaper describing the chameleon cloud service. Read Categorize

"Clouds and Reproducibility: A Way to Go to Scientific Experiments?" [193]
Cloud Computing: Principles, Systems and Applications (2017)
`cloud-computing-2017.pdf`

An article describing the role that cloud computing has to play in reproducible computation.   Read Categorize

"Towards Automating Data Narratives" [194]
Proceedings of the 22Nd International Conference on Intelligent User Interfaces (2017)
`p565-gil.pdf`

An article about data narratives. Read Categorize

Datasets for the journal Computer Physics Communications [195]
()

A repository of code and data for articles published in Computer Physics Communications.   Categorize

"Determinism and Reproducibility in Large-Scale HPC Systems" [196]
(2013)
`chiang-determinism-reproducibility-hpc-2013.pdf`

Article about determinism and reproducibility in HPC.   Read   Categorize

Standing Together for Reproducibility in Large-Scale Computing [197]
(2014)
`reproducibility_at_xsede_final_report_2014.pdf`

An overview of reproducibility at xsede14.   Read   Categorize

"Next-Generation Implications of Open Access" [198]
(2007)

A fun article 10 years ago that predicted widespread use of open coding.   Read   Categorize

"Reproducible Research Practices and Transparency across the Biomedical Literature" [199]
PLOS Biology (2016)
`iqbal-reproducible-bio-2016.PDF`

A study about reproducibility in biology.   Read   Categorize

"Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations" [200]
Journal of Clinical Epidemiology (XXXX)
`Ioannidis-vibration-epidemiology-2017.pdf`

A UQ study about model specification uncertainty.   Read   Categorize

"Anticipating consequences of sharing raw data and code and of awarding badges for sharing" [201]
Journal of Clinical Epidemiology (XXXX)
`Ioannidis-consequences-sharing-2017.pdf`

A paper about possible consequences from implementing policies which require or encourage code and data reproducibility and sharing.   Read   Categorize

"Facilitating the Reproducibility of Scientific Workflows with Execution Environment Specifications" [48]
Procedia Computer Science (2017)
`makeflow-2017.pdf`

Another scientific workflow system (from Thain).   Read   Categorize

o2r - opening reproducible research [61]
()

Another open scientific workflow platform.   Read   Categorize

"Software citation principles" [202]
PeerJ Computer Science (2016)
`peerj-cs-86.pdf`

An article about how to do software citation. (For Force11 citation working group)   Read   Categorize

Collected Algorithms of the ACM [62]
()

A collection of algorithms from the ACM. This is a long standing code repository.

"Reproducible research in signal processing" [203]
IEEE Signal Processing Magazine (2009)

Another possibly foundational reproducibility citation? (From [45]: Last but not least, recent studies indicate that reproducibility increases impact, visibility, and research quality.) Read   Categorize

"How scientists fool themselves – and how they can stop" [204]
Nature (2015)

Another good article about how reproducibility helps against scientists fooling themselves. (From [45]: and helps defeat self-deception)   Read   Categorize

"Statistical Analyses and Reproducible Research" [205]
Journal of Computational and Graphical Statistics (2007)
`gentleman-stat-analy-repro-2007.pdf`

This paper introduces the term 'compendium'. It defines the term to mean the collection of input data, source code, and any configuration parameters needed to replicate the published material. They also adopt a model-view-controller kind of scheme where the authors can create a so-called 'transformer' which takes the compendium output and creates necessary plots, tables, or images from said data which the author wants to convey.   Read More   Categorize

Banned By Gaussian [6]
()

A website dedicated to the ongoing scandal which Gaussian is perpetrating on the chemistry community. They decide to ban scientists who work on competing software thus blocking these scientists from being able to verify work published by other scientists using the Gaussian software.   Read   Categorize

"Guidelines for presentation of methodological choices in the publication of computational results. A. Ab initio electronic structure calculations (IUPAC Recommendations 1998)" [8]
Pure Appl. Chem. (1998)
`pac-guidelines-presentation-1998.pdf`

A set of guidelines from the Pure and Applied Chemistry Journal about the presentation of results in their journals. Circa 1998. Read Categorize

"Guidelines for publication of research results from force-field calculations(Technical Report)" [7]
Pure Appl. Chem. (1998)
`pac-guidelines-publication-1998.pdf`

A set of guidelines from the Pure and Applied Chemistry Journal about the publication of source code in their journals. Circa 1998. Referenced directly by [6], the relevant conclusions are: "2. Reproducibility of computations should be a defining goal in reporting computational results. Sufficient technical details should be provided to allow a reader to reproduce the results of any calculations. Note that strict reproducibility of computations may be precluded by differences among computers (e.g., differences in operating systems, host-based algorithms, machine precision, etc.), software (e.g., installation procedures and programming modifications), and applications (e.g., for probabilistic methods such as Monte Carlo calculations)." and "5. An adequate specification of the computer program used to carry out the calculations is essential. If the calculations were carried out with commercially available software, the investigator must list the source, version number, and force field employed. Any program modifications and any differences from the published version (including changes in parameters) must be fully described or available to other investigators upon request. Changes in parameters, program options (e.g., dielectric constant, nonbonded cutoffs, etc.) affecting reproducibility of the calculation, and program constraints must be fully described in the article. If the program is not available, commercially or otherwise, the authors must specify sufficient detail that the calculation could be reproduced." Read Categorize

"Guidelines for calibration in analytical chemistry. Part I. Fundamentals and single component calibration (IUPAC Recommendations 1998)" [206]
Pure Appl. Chem. (1998)
`pac-guidelines-calibration-1998.pdf`

A set of guidelines from the Pure and Applied Chemistry Journal about calibration. Read Categorize

"High drug attrition rates—where are we going wrong?" [207]
Nat Rev Clin Oncol (2011)
`hutchinson-drug-attrition-2011.pdf`

An Editorial reporting empirical measurements about clinical drug development in oncology. Read Categorize

"Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data" [208]
PLOS ONE (2010)
`fanelli-pressure-publish-2010.pdf`

An empirical study showing that pressure to publish can increase scientific bias. Read Categorize

"Hedging against academic risk" [209]
SciBX (2011)

```
osherovich-hedging-academic-2011.pdf
```

A discussion of the unwritten rule that 50% of all published academic works cannot be reproduced in an industrial lab. Read Categorize

"What errors do peer reviewers detect, and does training improve their ability to detect them?" [210]
Journal of the Royal Society of Medicine (2008)
```
schroter-reviewer-errors-2008.pdf
```

An article about reviewers and what errors they can and can't catch. And whether they can be trained to be better. "Reviewers found an average of 2.58 of 9 major errors with no notable difference between the groups. Interventions had small effects. Biased randomization was the error detected most frequently in these papers, with over 60% of reviewers rejeting the papers identifying this error. Reviewers who did not reject the papers found fewer errors and the proportion finding biased randomization was less than 40% for each paper." Read Categorize

"What happens to the manuscripts that have not been accepted for publication in Occupational and Environmental Medicine?" [211]
Occupational and Environmental Medicine (2001)
```
nemery-not-accepted-2001.pdf
```

An article discussing the fate of reserach which was rejected from Occupational and Environmental Medicine. 405 manuscripts were rejected (44%), and of those 218 (54%) were traced to 72 different journals. Read Categorize

"Fate of Submitted Manuscripts Rejected from the American Journal of Neuroradiology: Outcomes and Commentary" [212]
American Journal of Neuroradiology (2007)
```
mcdonald-fate-rejected-2007.pdf
```

An article discussing the fate of research rejected from the American Journal of Neruoradiology. "Of 554 rejected submissions, 315 (56%) were subsequently published in 115 different journals." There were more more detailed statistics. Read Categorize

"Trial watch: Phase II failures: 2008–2010" [22]
Nat Rev Drug Discov (2011)

An empirical article talking about phase II failures from 2008 to 2010. Used in [24] Don't currently have access to the article. Read Categorize

"Phase II and phase III failures: 2013-2015" [213]
Nat Rev Drug Discov (2016)

Another empirical article talking about phase II and II failures from 2013-2015. It seems to be a follow on to [22]. Don't currently have access to the article. Read Categorize

Open Data: the researcher perspective - survey and case studies [28]
(2017)

```
berghmans-survey-open-data-dataset-2017.pdf
```

A survey sent out to scientists. This citation has a bunch of statistics which need to be added to the empirical studies list.   Read   Categorize

"The poor availability of psychological research data for reanalysis." [21]
American Psychologist (2006)

This article describes the difficulty in obtaining data from published psychological studies. After more than 400 emails, and sending some corresponding authors detailed descriptions of their study aims, approvals by the ethical committee, and signed assurances not to share with others, they only ended up with 38 data sets from 64 studies (25.7% of the total number of 249 data sets).
We're required to purchase this article, so I can only read the abstract.   Read   Categorize

Program Logics for Certified Compilers [51]
 (2014)

A book describing the field of program logic verification and some tools that form the Verified Software Toolchain.   Read   Categorize

"Mathematics in the Age of the Turing Machine" [214]
 (2013)
```
hales-turing-machine-2013.pdf
```

An article reviewing mathematical proofs which rely on computers.   Read   Categorize

"Publish your computer code: it is good enough" [215]
Nature (2010)
```
barnes-code-good-enough-2010.pdf
```

Editorial describing some of the misconceptions and benefits of releasing source code.
Nice point: Code which isn't released, is sometimes deleted, thereby destroying a full record of what was done.
Interesting case study about NASA releasing climate code. People couldn't run it, an outsider cleaned it up, and now NASA considers replacing their code with the outsider's code.
Code not backed by experts: abandonware.
People will accept code which does what you say it does. If it has bugs you should release it, and if not, well you should probably fix it anyway.   Categorize

References

[1] *Artifact review and badging.* [Online]. Available: https://www.acm.org/publications/policies/artifact-review-badging (cit. on pp. 3, 37).

[2] *Neuromorphic chips are destined for deep learning—or obscurity.* [Online]. Available: http://spectrum.ieee.org/semiconductors/design/neuromorphic-chips-are-destined-for-deep-learningor-obscurity (cit. on pp. 4, 37).

[3]   N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, R. Boyle, P.-l. Cantin, C. Chao, C. Clark, J. Coriell, M. Daley, M. Dau, J. Dean, B. Gelb, T. V. Ghaemmaghami, R. Gottipati, W. Gulland, R. Hagmann, C. R. Ho, D. Hogberg, J. Hu, R. Hundt, D. Hurt, J. Ibarz, A. Jaffey, A. Jaworski, A. Kaplan, H. Khaitan, D. Killebrew, A. Koch, N. Kumar, S. Lacy, J. Laudon, J. Law, D. Le, C. Leary, Z. Liu, K. Lucke, A. Lundin, G. MacKean, A. Maggiore, M. Mahony, K. Miller, R. Nagarajan, R. Narayanaswami, R. Ni, K. Nix, T. Norrie, M. Omernick, N. Penukonda, A. Phelps, J. Ross, M. Ross, A. Salek, E. Samadiani, C. Severn, G. Sizikov, M. Snelham, J. Souter, D. Steinberg, A. Swing, M. Tan, G. Thorson, B. Tian, H. Toma, E. Tuttle, V. Vasudevan, R. Walter, W. Wang, E. Wilcox, and D. H. Yoon, "In-datacenter performance analysis of a tensor processing unit," in *Proceedings of the 44th Annual International Symposium on Computer Architecture*, ser. ISCA '17, Toronto, ON, Canada: ACM, 2017, pp. 1–12, ISBN: 978-1-4503-4892-8. DOI: 10.1145/3079856.3080246. [Online]. Available: http://doi.acm.org/10.1145/3079856.3080246 (cit. on pp. 4, 47).

[4]   D. H. Bailey, *Petascale computing and high-precision arithmetic: Applications and challenges*, Mar. 2015. [Online]. Available: http://www.davidhbailey.com/dhbtalks/dhb-zurich-hp.pdf (cit. on pp. 5, 8–10, 23, 25).

[5]   A. Marcus, "Pay up or retract? drug survey spurs conflict," *Science*, vol. 357, no. 6356, pp. 1085–1086, 2017, ISSN: 0036-8075. DOI: 10.1126/science.357.6356.1085. eprint: http://science.sciencemag.org/content/357/6356/1085.full.pdf. [Online]. Available: http://science.sciencemag.org/content/357/6356/1085 (cit. on pp. 5, 29).

[6]   *Banned by gaussian*. [Online]. Available: http://www.bannedbygaussian.org/ (cit. on pp. 6, 50, 51).

[7]   D. J. Raber and W. C. Guida, "Guidelines for publication of research results from force-field calculations(technical report)," *Pure Appl. Chem.*, vol. 70, no. 10, pp. 2047–2049, 1998. DOI: 10.1351/pac199870102047 (cit. on pp. 6, 22, 51).

[8]   J. E. Boggs, "Guidelines for presentation of methodological choices in the publication of computational results. a. ab initio electronic structure calculations (iupac recommendations 1998)," *Pure Appl. Chem.*, vol. 70, no. 4, pp. 1015–1018, 1998. DOI: 10.1351/pac199870041015 (cit. on pp. 6, 22, 51).

[9]   D. H. Bailey, "High-precision floating-point arithmetic in scientific computation," *Computing in Science Engineering*, vol. 7, no. 3, pp. 54–61, May 2005, ISSN: 1521-9615. DOI: 10.1109/MCSE.2005.52 (cit. on pp. 6, 8–10, 23, 25).

[10] "Ieee standard for floating-point arithmetic," *IEEE Std 754-2008*, pp. 1–70, Aug. 2008. DOI: 10.1109/IEEESTD.2008.4610935 (cit. on pp. 6–8, 39).

[11] "Ieee standard for floating-point arithmetic - redline," *IEEE Std 754-2008 (Revision of IEEE Std 754-1985) - Redline*, pp. 1–82, Aug. 2008. DOI: 10.1109/IEEESTD.2008.5976968 (cit. on pp. 6, 39).

[12] N. Higham, *Accuracy and Stability of Numerical Algorithms*, Second.: Society for Industrial and Applied Mathematics, 2002. DOI: 10.1137/1.9780898718027. eprint: http://epubs.siam.org/doi/pdf/10.1137/1.9780898718027. [Online]. Available: http://epubs.siam.org/doi/abs/10.1137/1.9780898718027 (cit. on pp. 7, 25).

[13] *The gnu c library: Rounding*. [Online]. Available: https://www.gnu.org/software/libc/manual/html%5C_node/Rounding.html (cit. on pp. 7, 25).

[14] D. H. Bailey, *Resolving numerical anomalies in scientific computation*, Feb. 2008. eprint: http://www.davidhbailey.com/dhbpapers/numerical-bugs.pdf. [Online]. Available: https://pubarchive.lbl.gov/islandora/object/ir%5C%3A150532 (cit. on pp. 7, 9, 10, 31).

[15] J. Demmel and H. D. Nguyen, "Fast reproducible floating-point summation," in *Computer Arithmetic (ARITH), 2013 21st IEEE Symposium on*, Apr. 2013, pp. 163–172. DOI: 10.1109/ARITH.2013.9 (cit. on pp. 9, 25).

[16] D. H. Bailey, *A thread-safe arbitrary precision computation package*. [Online]. Available: http://www.davidhbailey.com/dhbpapers/mpfun2015.pdf (cit. on pp. 9, 23, 34).

[17] C. Rubio-González, C. Nguyen, B. Mehne, K. Sen, J. Demmel, W. Kahan, C. Iancu, W. Lavrijsen, D. H. Bailey, and D. Hough, "Floating-point precision tuning using blame analysis," in *Proceedings of the 38th International Conference on Software Engineering*, ser. ICSE '16, Austin, Texas: ACM, 2016, pp. 1074–1085, ISBN: 978-1-4503-3900-1. DOI: 10.1145/2884781.2884850. [Online]. Available: http://doi.acm.org/10.1145/2884781.2884850 (cit. on pp. 10, 23, 34).

[18] W. G. Dewald, J. G. Thursby, and R. G. Anderson, "Replication in empirical economics: The journal of money, credit and banking project," *The American Economic Review*, vol. 76, no. 4, pp. 587–603, 1986, ISSN: 00028282. [Online]. Available: http://www.jstor.org/stable/1806061 (cit. on pp. 12, 24).

[19] L. Hatton and A. Roberts, "How accurate is scientific software?" *IEEE Transactions on Software Engineering*, vol. 20, no. 10, pp. 785–797, Oct. 1994, ISSN: 0098-5589. DOI: 10.1109/32.328993 (cit. on pp. 12, 25).

[20] R. D. Peng, F. Dominici, and S. L. Zeger, "Reproducible epidemiologic research," *American Journal of Epidemiology*, vol. 163, no. 9, pp. 783–789, 2006. DOI: 10.1093/aje/kwj093. eprint: /oup/backfile/content_public/journal/aje/163/9/10.1093/aje/kwj093/2/kwj093.pdf. [Online]. Available: http://dx.doi.org/10.1093/aje/kwj093 (cit. on pp. 12, 24).

[21] J. M. Wicherts, D. Borsboom, J. Kats, and D. Molenaar, "The poor availability of psychological research data for reanalysis.," *American Psychologist*, vol. 61, pp. 726–728, 7 Oct. 2006. DOI: 10.1037/0003-066X.61.7.726 (cit. on pp. 12, 53).

[22] J. Arrowsmith, "Trial watch: Phase ii failures: 2008–2010," *Nat Rev Drug Discov*, vol. 10, no. 5, pp. 328–329, May 2011, ISSN: 1474-1776. DOI: 10.1038/nrd3439. [Online]. Available: http://dx.doi.org/10.1038/nrd3439 (cit. on pp. 12, 52).

[23] F. Prinz, T. Schlange, and K. Asadullah, "Believe it or not: How much can we rely on published data on potential drug targets?" *Nat Rev Drug Discov*, vol. 10, no. 9, pp. 712–712, Sep. 2011, ISSN: 1474-1776. DOI: 10.1038/nrd3439-c1. [Online]. Available: http://dx.doi.org/10.1038/nrd3439-c1 (cit. on pp. 12, 24).

[24] C. G. Begley and L. M. Ellis, "Drug development: Raise standards for preclinical cancer research," *Nature*, vol. 483, no. 7391, pp. 531–533, "Mar" 2012, ISSN: 0028-0836. DOI: 10.1038/483531a. [Online]. Available: http://dx.doi.org/10.1038/483531a (cit. on pp. 12, 24, 52).

[25] A. C. Chang and P. Li, "Is economics research replicable? sixty published papers from thirteen journals say "usually not"," 2015. DOI: 10.17016/FEDS.2015.083 (cit. on pp. 12, 24).

[26] O. S. Collaboration, "Estimating the reproducibility of psychological science," *Science*, vol. 349, no. 6251, 2015, ISSN: 0036-8075. DOI: 10.1126/science.aac4716. eprint: http://science.sciencemag.org/content/349/6251/aac4716.full.pdf. [Online]. Available: http://science.sciencemag.org/content/349/6251/aac4716 (cit. on pp. 13, 27, 36).

[27] C. Collberg and T. A. Proebsting, "Repeatability in computer systems research," *Commun. ACM*, vol. 59, no. 3, pp. 62–69, Feb. 2016, ISSN: 0001-0782. DOI: 10.1145/2812803. [Online]. Available: http://doi.acm.org/10.1145/2812803 (cit. on pp. 13, 36).

[28] S. Berghmans, H. Cousijn, G. Deakin, I. Meijer, A. Mulligan, A. Plume, S. de Rijcke, A. Rushforth, C. Tatum, T. van Leeuwen, and L. Waltman, *Open data: The researcher perspective - survey and case studies*, Apr. 2017. DOI: 10.17632/bwrnfb4bvh.1 (cit. on pp. 13, 52).

[29] D. Bernstein, "Containers and cloud: From lxc to docker to kubernetes," vol. 1, no. 3, 2014, ISSN: 2325-6095. DOI: https://doi.org/10.1109/MCC.2014.51. [Online]. Available: http://ieeexplore.ieee.org/document/7036275/ (cit. on p. 14).

[30] S. McCarty, *Architecting containers part 1: Why understanding user space vs. kernel space matters*, Jul. 2015. [Online]. Available: http://rhelblog.redhat.com/2015/07/29/architecting-containers-part-1-user-space-vs-kernel-space/ (cit. on p. 14).

[31] Intel, "Linux containers streamline virtualization and complement hypervisor-based virtual machines," [Online]. Available: https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/linux-containers-hypervisor-based-vms-paper.pdf (cit. on p. 14).

[32] *Cgroups*. [Online]. Available: https://en.wikipedia.org/wiki/Cgroups (cit. on p. 15).

[33] P. Menage, *Cgroups*. [Online]. Available: https://www.kernel.org/doc/Documentation/cgroup-v1/cgroups.txt (cit. on p. 15).

[34] M. Kerrisk, "Namespaces in operation, part 1: Namespaces overview," Jan. 2013. [Online]. Available: https://lwn.net/Articles/531114/ (cit. on p. 15).

[35] *Lxc*. [Online]. Available: https://en.wikipedia.org/wiki/LXC#cite_note-lxc-1-0-security-features.2F-5 (cit. on p. 15).

[36] R. Hat, *Discretionary access control (dac)*. [Online]. Available: https://www.centos.org/docs/5/html/5.1/Deployment_Guide/sec-dac-intro1.html (cit. on p. 15).

[37] D. Merkel, "Docker: Lightweight linux containers for consistent development and deployment," *Linux J.*, vol. 2014, no. 239, Mar. 2014, ISSN: 1075-3583. [Online]. Available: http://dl.acm.org/citation.cfm?id=2600239.2600241 (cit. on pp. 16, 34).

[38] G. M. Kurtzer, V. Sochat, and M. W. Bauer, "Singularity: Scientific containers for mobility of compute," *PLOS ONE*, vol. 12, no. 5, pp. 1–20, May 2017. DOI: 10.1371/journal.pone.0177459. [Online]. Available: https://doi.org/10.1371/journal.pone.0177459 (cit. on pp. 16, 34).

[39] V. Stodden, *Data sharing and reproducibility*, Mar. 2017. [Online]. Available: http://stanford.edu/~vcs/talks/NASSackler2017-2-STODDEN.pdf (cit. on pp. 16, 46).

[40] I. Jimenez, M. Sevilla, N. Watkins, C. Maltzahn, J. Lofstead, K. Mohror, R. Arpaci-Dusseau, and A. Arpaci-Dusseau, "Standing on the shoulders of giants by managing scientific experiments like software," *;login: The USENIX Magazine*, vol. 41, no. 4, [Online]. Available: https://www.usenix.org/publications/login/winter2016/jimenez (cit. on pp. 17, 35).

[41] A. Ustyuzhanin, T. Head, I. Babuschkin, and A. Tiunov, "Everware toolkit. supporting reproducible science and challenge-driven education.," May 2017. eprint: https://arxiv.org/pdf/1703.01200.pdf (cit. on pp. 17, 45).

[42] N. Matloff, R. Davis, L. Beckett, and P. Thompson, "Revisit: A workflow tool for data science," *arXiv*, Aug. 2017. [Online]. Available: https://arxiv.org/abs/1708.04789 (cit. on pp. 17, 35).

[43] X. Han, T. Pasquier, T. Ranjan, M. Goldstein, and M. Seltzer, "Frappuccino: Fault-detection through runtime analysis of provenance," in *9th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 17)*, Santa Clara, CA: USENIX Association, 2017. [Online]. Available: https://www.usenix.org/conference/hotcloud17/program/presentation/han (cit. on pp. 17, 38).

[44] P. Thavasimani and P. Missier, "Facilitating reproducible research by investigating computational metadata," in *2016 IEEE International Conference on Big Data (Big Data)*, Dec. 2016, pp. 3045–3051. DOI: 10.1109/BigData.2016.7840958 (cit. on pp. 17, 38).

[45] F. Chirigati, R. Rampin, D. Shasha, and J. Freire, "Reprozip: Computational reproducibility with ease," in *Proceedings of the 2016 International Conference on Management of Data*, ser. SIGMOD '16, San Francisco, California, USA: ACM, 2016, pp. 2085–2088, ISBN: 978-1-4503-3531-7. DOI: 10.1145/2882903.2899401. [Online]. Available: http://doi.acm.org/10.1145/2882903.2899401 (cit. on pp. 18, 19, 26, 33, 50).

[46] P. Ivie and D. Thain, "Prune: A preserving run environment for reproducible scientific computing," in *2016 IEEE 12th International Conference on e-Science (e-Science)*, Oct. 2016, pp. 61–70. DOI: 10.1109/eScience.2016.7870886 (cit. on pp. 18, 34).

[47] H. Meng, D. Thain, A. Vyushkov, M. Wolf, and A. Woodard, "Conducting reproducible research with umbrella: Tracking, creating, and preserving execution environments," in *2016 IEEE 12th International Conference on e-Science (e-Science)*, Oct. 2016, pp. 91–100. DOI: 10.1109/eScience.2016.7870889 (cit. on pp. 18, 44).

[48] H. Meng and D. Thain, "Facilitating the reproducibility of scientific workflows with execution environment specifications," *Procedia Computer Science*, vol. 108, pp. 705–714, 2017, International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland, ISSN: 1877-0509. DOI: http://dx.doi.org/10.1016/j.procs.2017.05.116. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050917306816 (cit. on pp. 18, 49).

[49] T. Gamblin, M. LeGendre, M. R. Collette, G. L. Lee, A. Moody, B. R. de Supinski, and S. Futral, "The spack package manager: Bringing order to hpc software chaos," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '15, Austin, Texas: ACM, 2015, 40:1–40:12, ISBN: 978-1-4503-3723-6. DOI: 10.1145/2807591.2807623. [Online]. Available: http://doi.acm.org/10.1145/2807591.2807623 (cit. on pp. 19, 32).

[50] K. Hoste, J. Timmerman, A. Georges, and S. D. Weirdt, "Easybuild: Building software with ease," in *2012 SC Companion: High Performance Computing, Networking Storage and Analysis*, Nov. 2012, pp. 572–582. DOI: 10.1109/SC.Companion.2012.81 (cit. on p. 19).

[51] A. W. Appel, *Program Logics for Certified Compilers*. Cambridge University Press, 2014, ISBN: 9781107048010 (cit. on pp. 19, 53).

[52] *Reproducible-builds.org*. [Online]. Available: https://reproducible-builds.org/ (cit. on pp. 20, 33).

[53] D. H. T. That, G. Fils, Z. Yuan, and T. Malik, "Sciunits: Reusable research objects," 2017. [Online]. Available: https://arxiv.org/abs/1707.05731 (cit. on pp. 20, 33).

[54] *Computer physics communications*. [Online]. Available: https://www.journals.elsevier.com/computer-physics-communications/ (cit. on pp. 20, 33).

[55] *Darpa open catalog*. [Online]. Available: https://opencatalog.darpa.mil/ (cit. on pp. 20, 33).

[56] *Doe code*. [Online]. Available: https://www.osti.gov/doecode/ (cit. on pp. 20, 33).

[57] *Mendeley data repository for comptuter physics communications*. [Online]. Available: https://data.mendeley.com/datasets/journals/00104655 (cit. on pp. 20, 34).

[58] *Ipol journal - image processing on line*. [Online]. Available: http://www.ipol.im/ (cit. on pp. 20, 41).

[59] J. L. Ayala, "Code ocean is live: Upload your algorithms," *IEEE Design Test*, vol. 34, no. 3, pp. 108–109, Jun. 2017, ISSN: 2168-2356. DOI: 10.1109/MDAT.2017.2692207 (cit. on pp. 20, 41).

[60] J. Mambretti, J. Chen, and F. Yeh, "Next generation clouds, the chameleon cloud testbed, and software defined networking (sdn)," in *Proceedings of the 2015 International Conference on Cloud Computing Research and Innovation (ICCCRI)*, ser. ICCCRI '15, Washington, DC, USA: IEEE Computer Society, 2015, pp. 73–79, ISBN: 978-1-5090-0144-6. DOI: 10.1109/ICCCRI.2015.10. [Online]. Available: http://dx.doi.org/10.1109/ICCCRI.2015.10 (cit. on pp. 20, 48).

[61] *O2r - opening reproducible research*. [Online]. Available: https://o2r.uni-muenster.de/ (cit. on pp. 20, 50).

[62] *Collected algorithms of the acm.* [Online]. Available: http://netlib.org/toms/ (cit. on pp. 20, 50).

[63] D. T. Brown, *Adventures in replicable scientific papers: Docker*, May 2015. [Online]. Available: http://ivory.idyll.org/blog/2015-docker-and-replicating-papers.html (cit. on pp. 21, 42).

[U.S. Const.] Constitution of the united states, 1787 (cit. on p. 21).

[64] *ACM SIGHPC Connect*, vol. 5, no. 1, R. Vuduc, Ed., 2016 (cit. on pp. 21, 29).

[65] *ACM SIGHPC Connect*, vol. 5, no. 2, R. Vuduc, Ed., 2017 (cit. on pp. 21, 29).

[66] R. F. Boisvert, "Incentivizing reproducibility," *Commun. ACM*, vol. 59, no. 10, pp. 5–5, Sep. 2016, ISSN: 0001-0782. DOI: 10.1145/2994031. [Online]. Available: http://doi.acm.org/10.1145/2994031 (cit. on pp. 21, 29).

[67] K. Sandler, L. Ohrstrom, L. Moy, and R. McVay, "Killed by code: Software transparency in implantable medical devices," *Software Freedom Law Center*, Accessed: 2017-08-28. [Online]. Available: https://www.softwarefreedom.org/resources/2010/transparent-medical-devices.pdf (cit. on pp. 22, 41).

[68] D. F. Bauer, "Constructing confidence sets using rank statistics," *Journal of the American Statistical Association*, vol. 67, no. 339, pp. 687–690, 1972. DOI: 10.1080/01621459.1972.10481279. eprint: http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1972.10481279. [Online]. Available: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1972.10481279 (cit. on pp. 22, 38).

[69] *Wilcoxon rank sum and signed rank tests.* [Online]. Available: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/wilcox.test.html (cit. on pp. 22, 38).

[70] A. C. A. Hope, "A simplified monte carlo significance test procedure," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 30, no. 3, pp. 582–598, 1968, ISSN: 00359246. [Online]. Available: http://www.jstor.org/stable/2984263 (cit. on pp. 22, 39).

[71] *Pearson's chi-squared test for count data.* [Online]. Available: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/chisq.test.html (cit. on pp. 22, 39).

[72] V. Stodden, "The legal framework for reproducible scientific research: Licensing and copyright," *Computing in Science Engineering*, vol. 11, no. 1, pp. 35–40, Jan. 2009, ISSN: 1521-9615. DOI: 10.1109/MCSE.2009.19 (cit. on pp. 23, 29).

[73] D. L. Donoho, A. Maleki, I. U. Rahman, M. Shahram, and V. Stodden, "Reproducible research in computational harmonic analysis," *Computing in Science & Engineering*, vol. 11, no. 1, pp. 8–18, 2009. DOI: 10.1109/MCSE.2009.15. eprint: http://aip.scitation.org/doi/pdf/10.1109/MCSE.2009.15. [Online]. Available: http://aip.scitation.org/doi/abs/10.1109/MCSE.2009.15 (cit. on pp. 23, 26).

[74] D. H. Bailey, *David h bailey.* [Online]. Available: http://davidhbailey.com/ (cit. on p. 23).

[75] C. Collberg, T. Proebsting, G. Moraila, A. Shankaran, Z. Shi, and A. M. Warren, "Measuring reproducibility in computer systems research," *Unpublished report*, Oct. 2014. [Online]. Available: https://reproducibility.cs.arizona.edu (cit. on p. 26).

[76] D. Ince, "The problem of reproducibility," *CHANCE*, vol. 25, no. 3, pp. 4–7, 2012. DOI: 10.1080/09332480.2012.726554. eprint: http://dx.doi.org/10.1080/09332480.2012.726554. [Online]. Available: http://dx.doi.org/10.1080/09332480.2012.726554 (cit. on p. 26).

[77] J. Gustedt, E. Jeannot, and M. Quinson, "Experimental Methodologies for Large-Scale Systems: a Survey," *Parallel Processing Letters*, vol. 19, no. 3, pp. 399–418, 2009. DOI: 10.1142/S0129626409000304. [Online]. Available: https://hal.inria.fr/inria-00364180 (cit. on p. 26).

[78] D. Garijo, S. Kinnings, L. Xie, L. Xie, Y. Zhang, P. E. Bourne, and Y. Gil, "Quantifying reproducibility in computational biology: The case of the tuberculosis drugome," *PLoS ONE*, vol. 8, no. 11, pp. 1–11, Nov. 2013. DOI: 10.1371/journal.pone.0080278. [Online]. Available: http://dx.doi.org/10.1371%5C%2Fjournal.pone.0080278 (cit. on p. 26).

[79] F. Chirigati, R. Capone, R. Rampin, J. Freire, and D. Shasha, "A collaborative approach to computational reproducibility," *Information Systems*, vol. 59, pp. 95–97, 2016, ISSN: 0306-4379. DOI: http://dx.doi.org/10.1016/j.is.2016.03.002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306437916300813 (cit. on p. 26).

[80] S. R. Piccolo and M. B. Frampton, "Tools and techniques for computational reproducibility," *GigaScience*, vol. 5, no. 1, pp. 1–13, 2016, ISSN: 2047-217X. DOI: 10.1186/s13742-016-0135-4. [Online]. Available: http://dx.doi.org/10.1186/s13742-016-0135-4 (cit. on p. 26).

[81] J. P. A. Ioannidis, D. B. Allison, C. A. Ball, I. Coulibaly, X. Cui, A. C. Culhane, M. Falchi, C. Furlanello, L. Game, G. Jurman, J. Mangion, T. Mehta, M. Nitzberg, G. P. Page, E. Petretto, and V. van Noort, "Repeatability of published microarray gene expression analyses," *Nat Genet*, vol. 41, no. 2, pp. 149–155, Feb. 2009, ISSN: 1061-4036. DOI: 10.1038/ng.295. [Online]. Available: http://dx.doi.org/10.1038/ng.295 (cit. on p. 27).

[82] A. Eklund, T. E. Nichols, and H. Knutsson, "Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates," *Proceedings of the National Academy of Sciences*, vol. 113, no. 28, pp. 7900–7905, 2016. DOI: 10.1073/pnas.1602413113. eprint: http://www.pnas.org/content/113/28/7900.full.pdf. [Online]. Available: http://www.pnas.org/content/113/28/7900.abstract (cit. on p. 27).

[83] J. Lewis, C. E. Breeze, J. Charlesworth, O. J. Maclaren, and J. Cooper, "Where next for the reproducibility agenda in computational biology?" *BMC Systems Biology*, vol. 10, no. 1, pp. 1–10, 2016, ISSN: 1752-0509. DOI: 10.1186/s12918-016-0288-x. [Online]. Available: http://dx.doi.org/10.1186/s12918-016-0288-x (cit. on p. 27).

[84] P. Ivie, C. Zheng, and D. Thain, "A first look at reproducibility and non-determinism in cms software and root data," *CMS Technical Report TR-2016-01*, Oct. 2016 (cit. on p. 27).

[85] S. A., "Software simplified," *Nature*, vol. 546, pp. 173–174, May 2017. DOI: 10.1038/546173a. [Online]. Available: http://dx.doi.org/10.1038/546173a (cit. on p. 27).

[86] E. H. B. M. Gronenschild, P. Habets, H. I. L. Jacobs, R. Mengelers, N. Rozendaal, J. van Os, and M. Marcelis, "The effects of freesurfer version, workstation type, and macintosh operating system version on anatomical volume and cortical thickness measurements," *PLOS ONE*, vol. 7, no. 6, pp. 1–13, Jun. 2012. DOI: 10.1371/journal.pone.0038234. [Online]. Available: https://doi.org/10.1371/journal.pone.0038234 (cit. on p. 27).

[87] R. F. Service, "Keeping the faith," *Science*, vol. 357, no. 6356, pp. 1088–1091, 2017, ISSN: 0036-8075. DOI: 10.1126/science.357.6356.1088. eprint: http://science.sciencemag.org/content/357/6356/1088.full.pdf. [Online]. Available: http://science.sciencemag.org/content/357/6356/1088 (cit. on p. 27).

[88] T. Hoefler and R. Belli, "Scientific benchmarking of parallel computing systems: Twelve ways to tell the masses when reporting performance results," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '15, Austin, Texas: ACM, 2015, 73:1–73:12, ISBN: 978-1-4503-3723-6. DOI: 10.1145/2807591.2807644. [Online]. Available: http://doi.acm.org/10.1145/2807591.2807644 (cit. on p. 28).

[89] W. Gropp, E. Lusk, N. Doss, and A. Skjellum, "A high-performance, portable implementation of the mpi message passing interface standard," *Parallel Computing*, vol. 22, no. 6, pp. 789–828, 1996, ISSN: 0167-8191. DOI: http://dx.doi.org/10.1016/0167-8191(96)00024-5. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0167819196000245 (cit. on p. 28).

[90] R. Thakur, E. Lusk, and W. Gropp, "I/o in parallel applications: The weakest link," *International Journal of High Performance Computing Applications*, vol. 12, no. 4, pp. 389–395, 1998. DOI: 10.1177/109434209801200401. eprint: http://hpc.sagepub.com/content/12/4/389.full.pdf+html. [Online]. Available: http://hpc.sagepub.com/content/12/4/389.abstract (cit. on p. 28).

[91] R. Thakur and W. Gropp, "Test suite for evaluating performance of multithreaded {mpi} communication," *Parallel Computing*, vol. 35, no. 12, pp. 608–617, 2009, Selected papers from the 14th European PVM/MPI Users Group Meeting, ISSN: 0167-8191. DOI: http://dx.doi.org/10.1016/j.parco.2008.12.013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167819109000143 (cit. on p. 28).

[92] W. Gropp and E. L. Lusk, "Reproducible measurements of mpi performance characteristics," in *Proceedings of the 6th European PVM/MPI Users' Group Meeting on Recent Advances in Parallel Virtual Machine and Message Passing Interface*, London, UK, UK: Springer-Verlag, 1999, pp. 11–18, ISBN: 3-540-66549-8. [Online]. Available: http://dl.acm.org/citation.cfm?id=648136.748782 (cit. on p. 28).

[93] G. King, "Replication, replication," *PS: Political Science and Politics*, vol. 28, pp. 444–452, Sep. 1995, See updates to this paper for how I use this paper as a class assignment now. (cit. on p. 29).

[94] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Dułak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iuşan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Küçükbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G.-X. Zhang, and S. Cottenier, "Reproducibility in density functional theory calculations of solids," *Science*, vol. 351, no. 6280, 2016, ISSN: 0036-8075. DOI: 10.1126/science.aad3000. eprint: http://science.sciencemag.org/content/351/6280/aad3000.full.pdf. [Online]. Available: http://science.sciencemag.org/content/351/6280/aad3000 (cit. on p. 30).

[95] J. J. Wilke, K. Sargsyan, J. P. Kenny, B. Debusschere, H. N. Najm, and G. Hendry, "Validation and uncertainty assessment of extreme-scale hpc simulation through bayesian inference," in *Euro-Par 2013 Parallel Processing: 19th International Conference, Aachen, Germany, August 26-30, 2013. Proceedings*, F. Wolf, B. Mohr, and D. an Mey, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 41–52, ISBN: 978-3-642-40047-6. DOI: 10.1007/978-3-642-40047-6_7. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-40047-6%5C_7 (cit. on p. 30).

[96] P. Pavlidis, J. D. Jensen, W. Stephan, and A. Stamatakis, "A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans," *Molecular Biology and Evolution*, vol. 29, no. 10, p. 3237, 2012. DOI: 10.1093/molbev/mss136. eprint: /oup/backfile/Content_public/Journal/mbe/29/10/10.1093/molbev/mss136/2/mss136.pdf. [Online]. Available: +%20http://dx.doi.org/10.1093/molbev/mss136 (cit. on p. 30).

[97] A. Fokkens, M. Erp, M. Postma, T. Pedersen, P. Vossen, and N. Freire, "Offspring from reproduction problems: What replication failure teaches us," [Online]. Available: https://aclweb.org/anthology/P/P13/P13-1166.pdf (cit. on p. 30).

[98] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statist. Sci.*, vol. 4, no. 4, pp. 409–423, Nov. 1989. DOI: 10.1214/ss/1177012413. [Online]. Available: http://dx.doi.org/10.1214/ss/1177012413 (cit. on pp. 30, 45).

[99] C. D. Martino, W. Kramer, Z. Kalbarczyk, and R. Iyer, "Measuring and understanding extreme-scale application resilience: A field study of 5,000,000 hpc application runs," in *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, Jun. 2015, pp. 25–36. DOI: 10.1109/DSN.2015.50 (cit. on p. 31).

[100] T. Flouri, K. Kobert, T. Rognes, and A. Stamatakis, "Are all global alignment algorithms and implementations correct?" *bioRxiv*, 2015. DOI: 10.1101/031500. eprint: http://biorxiv.org/content/early/2015/11/12/031500.full.pdf. [Online]. Available: http://biorxiv.org/content/early/2015/11/12/031500 (cit. on p. 31).

[101] D. Darriba, T. Flouri, and A. Stamatakis, "The state of software in evolutionary biology," *bioRxiv*, 2015. DOI: 10.1101/031930. eprint: http://biorxiv.org/content/early/2015/11/16/031930.full.pdf. [Online]. Available: http://biorxiv.org/content/early/2015/11/16/031930 (cit. on p. 31).

[102] L. Hatton, "High-integrity software, computation and the scientific method," 2011 (cit. on p. 31).

[103] D. Monniaux, "The pitfalls of verifying floating-point computations," *ACM Trans. Program. Lang. Syst.*, vol. 30, no. 3, 12:1–12:41, May 2008, ISSN: 0164-0925. DOI: 10.1145/1353445.1353446. [Online]. Available: http://doi.acm.org/10.1145/1353445.1353446 (cit. on p. 31).

[104] S. McConnell, *Code Complete*. 2004 (cit. on p. 31).

[105] *Software engineering – product quality*, Standard, Mar. 2011 (cit. on p. 32).

[106] L. Hatton, "The chimera of software quality," *Computer*, vol. 40, no. 8, pp. 104–103, Aug. 2007, ISSN: 0018-9162. DOI: 10.1109/MC.2007.292 (cit. on p. 32).

[107] M. P. Forum, "Mpi: A message-passing interface standard," Knoxville, TN, USA, Tech. Rep., 1994 (cit. on p. 32).

[108] L. Dagum and R. Menon, "Openmp: An industry-standard api for shared-memory programming," *IEEE Comput. Sci. Eng.*, vol. 5, no. 1, pp. 46–55, Jan. 1998, ISSN: 1070-9924. DOI: 10.1109/99.660313. [Online]. Available: http://dx.doi.org/10.1109/99.660313 (cit. on p. 32).

[109] G. L. Bryan, M. L. Norman, B. W. O'Shea, T. Abel, J. H. Wise, M. J. Turk, D. R. Reynolds, D. C. Collins, P. Wang, S. W. Skillman, B. Smith, R. P. Harkness, J. Bordner, J.-h. Kim, M. Kuhlen, H. Xu, N. Goldbaum, C. Hummels, A. G. Kritsuk, E. Tasker, S. Skory, C. M. Simpson, O. Hahn, J. S. Oishi, G. C. So, F. Zhao, R. Cen, Y. Li, and The Enzo Collaboration, "ENZO: An Adaptive Mesh Refinement Code for Astrophysics," ApJS, vol. 211, 19, p. 19, Apr. 2014. DOI: 10.1088/0067-0049/211/2/19. arXiv: 1307.2265 [astro-ph.IM] (cit. on p. 32).

[110] P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu, "The rockstar phase-space temporal halo finder and the velocity offsets of cluster cores," *The Astrophysical Journal*, vol. 762, no. 2, p. 109, 2013. [Online]. Available: http://stacks.iop.org/0004-637X/762/i=2/a=109 (cit. on p. 32).

[111] The HDF Group. (2000-2010). Hierarchical data format version 5, [Online]. Available: http://www.hdfgroup.org/HDF5 (cit. on p. 33).

[112] M. J. Turk, B. D. Smith, J. S. Oishi, S. Skory, S. W. Skillman, T. Abel, and M. L. Norman, "yt: A Multi-code Analysis Toolkit for Astrophysical Simulation Data," ApJS, vol. 192, 9, p. 9, Jan. 2011. DOI: 10.1088/0067-0049/192/1/9. arXiv: 1011.3514 [astro-ph.IM] (cit. on p. 33).

[113] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing, "Jupyter notebooks – a publishing format for reproducible computational workflows," in *Proceedings of the 20th International Conference on Electronic Publishing*. 2016, pp. 87–90. DOI: 10.3233/978-1-61499-649-1-87 (cit. on p. 34).

[114] R Development Core Team, *R: A language and environment for statistical computing*, ISBN 3-900051-07-0, R Foundation for Statistical Computing, Vienna, Austria, 2008. [Online]. Available: http://www.R-project.org (cit. on p. 34).

[115] L. Czech, J. Huerta-Cepas, and A. Stamatakis, "A critical review on the use of support values in tree viewers and bioinformatics toolkits," *bioRxiv*, 2017. DOI: 10.1101/035360. eprint: http://biorxiv.org/content/early/2017/01/13/035360.full.pdf. [Online]. Available: http://biorxiv.org/content/early/2017/01/13/035360 (cit. on p. 35).

[116] L. Hatton and G. Warr, "Full computational reproducibility in biological science: Methods, software and a case study in protein biology," 2016. [Online]. Available: https://arxiv.org/abs/1608.06897 (cit. on p. 35).

[117] J.-h. Kim, O. Agertz, R. Teyssier, M. J. Butler, D. Ceverino, J.-H. Choi, R. Feldmann, B. W. Keller, A. Lupi, T. Quinn, Y. Revaz, S. Wallace, N. Y. Gnedin, S. N. Leitner, S. Shen, B. D. Smith, R. Thompson, M. J. Turk, T. Abel, K. S. Arraki, S. M. Benincasa, S. Chakrabarti, C. DeGraf, A. Dekel, N. J. Goldbaum, P. F. Hopkins, C. B. Hummels, A. Klypin, H. Li, P. Madau, N. Mandelker, L. Mayer, K. Nagamine, S. Nickerson, B. W. O'Shea, J. R. Primack, S. Roca-Fàbrega, V. Semenov, I. Shimizu, C. M. Simpson, K. Todoroki, J. W. Wadsley, J. H. Wise, and A. Collaboration, "The agora high-resolution galaxy simulations comparison project. ii. isolated disk test," *The Astrophysical Journal*, vol. 833, no. 2, p. 202, 2016. [Online]. Available: http://stacks.iop.org/0004-637X/833/i=2/a=202 (cit. on p. 35).

[118] V. Stodden, M. McNutt, D. H. Bailey, E. Deelman, Y. Gil, B. Hanson, M. A. Heroux, J. P. Ioannidis, and M. Taufer, "Enhancing reproducibility for computational methods," *Science*, vol. 354, no. 6317, pp. 1240–1241, 2016, ISSN: 0036-8075. DOI: 10.1126/science.aah6168. eprint: http://science.sciencemag.org/content/354/6317/1240.full.pdf. [Online]. Available: http://science.sciencemag.org/content/354/6317/1240 (cit. on p. 35).

[119] V. Stodden, "What computational scientists need to know about intellectual property law: A primer," in *Implementing Reproducible Research*, Chapman and Hall/CRC, Apr. 2014. [Online]. Available: https://web.stanford.edu/~vcs/papers/IRR-Stodden%5C_chapter.pdf (cit. on p. 36).

[120] H. Monajemi, D. L. Donoho, and V. Stodden, "Making massive computational experiments painless," in *2016 IEEE International Conference on Big Data (Big Data)*, Dec. 2016, pp. 2368–2373. DOI: 10.1109/BigData.2016.7840870 (cit. on p. 36).

[121] I. JA, "Contradicted and initially stronger effects in highly cited clinical research," *JAMA*, vol. 294, no. 2, pp. 218–228, 2005. DOI: 10.1001/jama.294.2.218. eprint: /data/journals/jama/4983/joc50060.pdf. [Online]. Available: +%20http://dx.doi.org/10.1001/jama.294.2.218 (cit. on p. 36).

[122] J. P. A. Ioannidis, "Why most published research findings are false," *PLOS Medicine*, vol. 2, no. 8, Aug. 2005. DOI: 10.1371/journal.pmed.0020124. [Online]. Available: https://doi.org/10.1371/journal.pmed.0020124 (cit. on p. 36).

[123] M. Baker, "1,500 scientists lift the lid on reproducibility," *Nature*, vol. 533, pp. 452–454, May 2016. DOI: 10.1038/533452a (cit. on p. 36).

[124] C. G. Begley and J. P. Ioannidis, "Reproducibility in science," *Circulation Research*, vol. 116, no. 1, pp. 116–126, 2014, ISSN: 0009-7330. DOI: 10.1161/CIRCRESAHA.114.303819. eprint: http://circres.ahajournals.org/content/116/1/116.full.pdf. [Online]. Available: http://circres.ahajournals.org/content/116/1/116 (cit. on p. 36).

[125] V. Stodden, *How computational science is changing the scientific method*, Jul. 2009. [Online]. Available: https://web.stanford.edu/~vcs/talks/Science20July2009VictoriaStodden.pdf (cit. on p. 36).

[126] J. B. Buckheit and D. L. Donoho, "Wavelab and reproducible research," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York, NY: Springer New York, 1995, pp. 55–81, ISBN: 978-1-4612-2544-7. DOI: 10.1007/978-1-4612-2544-7_5. [Online]. Available: https://doi.org/10.1007/978-1-4612-2544-7%5C_5 (cit. on p. 37).

[127] *Reproducible computational research: A history of hurdles, mostly overcome.* [Online]. Available: http://sepwww.stanford.edu/sep/jon/reproducible.html (cit. on p. 37).

[128] J. F. Claerbout and M. Karrenbach, "Electronic documents give reproducible research a new meaning," in *SEG Technical Program Expanded Abstracts*, 1992, pp. 601–604. DOI: 10.1190/1.1822162. [Online]. Available: http://sepwww.stanford.edu/doku.php?id=sep:research:reproducible:seg92 (cit. on p. 37).

[129] S. Fomel and J. F. Claerbout, "Reproducible research," *Computing in Science & Engineering*, vol. 11, no. 1, pp. 5–7, 2009. DOI: 10.1109/MCSE.2009.14. eprint: http://aip.scitation.org/doi/pdf/10.1109/MCSE.2009.14. [Online]. Available: http://aip.scitation.org/doi/abs/10.1109/MCSE.2009.14 (cit. on p. 37).

[130] M. Gavish and D. Donoho, "Three dream applications of verifiable computational results," *Computing in Science Engineering*, vol. 14, no. 4, pp. 26–31, Jul. 2012, ISSN: 1521-9615. DOI: 10.1109/MCSE.2012.65 (cit. on p. 37).

[131] S. Pino, L. Pollock, and S. Chandrasekaran, "Exploring translation of openmp to openacc 2.5: Lessons learned," in *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, May 2017, pp. 673–682. DOI: 10.1109/IPDPSW.2017.84 (cit. on pp. 37, 38).

[132] P. Ginsparg, "Next-generation implications of open access," *CTWatch Quarterly*, vol. 3, no. 3, Aug. 2007. [Online]. Available: http://www.ctwatch.org/quarterly/articles/2007/08/next-generation-implications-of-open-access/ (cit. on p. 38).

[133] V. Stodden, "The scientific method in practice: Reproducibility in the computational sciences," *MIT Sloan Research Paper*, no. 4773-10, Feb. 2010. DOI: 10.2139/ssrn.1550193. [Online]. Available: https://ssrn.com/abstract=1550193.37 (cit. on p. 39).

[134] M. Taufer, "Sc16 explores reproducibility for advanced computing through student cluster competition," [Online]. Available: http://sc16.supercomputing.org/2016/03/16/sc16-explores-reproducibility-advanced-computing-student-competition-michela-taufer/index.html (cit. on p. 39).

[135] "Facilitating reproducibility in scientific computing: Principles and practice," in *Reproducibility*. John Wiley & Sons, Inc., 2016, pp. 205–231, ISBN: 9781118865064. DOI: 10.1002/9781118865064.ch9. [Online]. Available: http://dx.doi.org/10.1002/9781118865064.ch9 (cit. on p. 39).

[136] R. Hanisch, M. Hildreth, L. McEwen, V. Stodden, G. Watts, D. S. Katz, N. Meyers, and A. E. Sands, *Workshops to gauge the impact of requirements for public access to data produced by nsf-funded research in mathematics and the physical sciences*, 2016. [Online]. Available: https://mpsopendata.crc.nd.edu/images/Reports/MPS%5C_Workshop%5C_Report%5C_FinalVersion.pdf (cit. on p. 39).

[137] *Performance development.* [Online]. Available: https://www.top500.org/statistics/perfdevel/ (cit. on p. 40).

[138] R. Cowen, "No evidence for or against gravitational waves," *Nature*, May 2014. DOI: 10.1038/nature.2014.15322 (cit. on p. 40).

[139] M. J. Mortonson and U. Seljak, "A joint analysis of planck and bicep2 b modes including dust polarization uncertainty," *Journal of Cosmology and Astroparticle Physics*, vol. 2014, no. 10, p. 035, 2014. [Online]. Available: http://stacks.iop.org/1475-7516/2014/i=10/a=035 (cit. on p. 40).

[140] R. Flauger, J. C. Hill, and D. N. Spergel, "Toward an understanding of foreground emission in the bicep2 region," *Journal of Cosmology and Astroparticle Physics*, vol. 2014, no. 08, p. 039, 2014. [Online]. Available: http://stacks.iop.org/1475-7516/2014/i=08/a=039 (cit. on p. 40).

[141] D. Bailey, R. Barrio, and J. Borwein, "High-precision computation: Mathematical physics and dynamics," *Applied Mathematics and Computation*, vol. 218, no. 20, pp. 10 106–10 121, 2012, ISSN: 0096-3003. DOI: http://dx.doi.org/10.1016/j.amc.2012.03.087. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0096300312003505 (cit. on p. 40).

[142] D. H. Bailey and J. M. Borwein, "High-precision arithmetic in mathematical physics," *Mathematics*, vol. 3, no. 2, pp. 337–367, 2015, ISSN: 2227-7390. DOI: 10.3390/math3020337. [Online]. Available: http://www.mdpi.com/2227-7390/3/2/337 (cit. on p. 40).

[143] C. Rubio-González, C. Nguyen, H. D. Nguyen, J. Demmel, W. Kahan, K. Sen, D. H. Bailey, C. Iancu, and D. Hough, "Precimonious: Tuning assistant for floating-point precision," in *2013 SC - International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, Nov. 2013, pp. 1–12. DOI: 10.1145/2503210.2503296 (cit. on p. 40).

[144] V. Stodden, D. H. Bailey, J. Borswein, R. J. LeVeque, W. Rider, and W. Stein, "Setting the default to reproducible," *ICERM Workshop 2013 Report*, eprint: http://www.davidhbailey.com/dhbpapers/icerm-report.pdf. [Online]. Available: https://www.carma.newcastle.edu.au/jon/icerm12.pdf (cit. on p. 40).

[145] *Scandal rocks scientific community*, Accessed: 2017-07-14. [Online]. Available: http://www.dw.com/en/scandal-rocks-scientific-community/a-646321 (cit. on p. 41).

[146] D. Bailey, R. Barrio, and J. Borwein, "High-precision computation: Mathematical physics and dynamics," *Applied Mathematics and Computation*, vol. 218, no. 20, pp. 10 106–10 121, 2012, ISSN: 0096-3003. DOI: http://dx.doi.org/10.1016/j.amc.2012.03.087. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0096300312003505 (cit. on p. 41).

[147] A. Bazavov, D. Toussaint, C. Bernard, J. Laiho, B. Billeter, C. DeTar, L. Levkova, M. B. Oktay, S. Gottlieb, U. M. Heller, J. E. Hetrick, J. Osborn, R. L. Sugar, and R. S. Van de Water, "Topological susceptibility with the asqtad action," *Phys. Rev. D*, vol. 81, p. 114 501, 11 Jun. 2010. DOI: 10.1103/PhysRevD.81.114501. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.81.114501 (cit. on p. 41).

[148] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn Reson Med*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007 (cit. on p. 41).

[149] M. A. Heroux, "Editorial: Acm toms replicated computational results initiative," *ACM Trans. Math. Softw.*, vol. 41, no. 3, 13:1–13:5, Jun. 2015, ISSN: 0098-3500. DOI: 10.1145/2743015. [Online]. Available: http://doi.acm.org/10.1145/2743015 (cit. on p. 41).

[150] S. Edwards, X. Liu, and N. Riga, "Creating repeatable computer science and networking experiments on shared, public testbeds," *SIGOPS Oper. Syst. Rev.*, vol. 49, no. 1, pp. 90–99, Jan. 2015, ISSN: 0163-5980. DOI: 10.1145/2723872.2723884. [Online]. Available: http://doi.acm.org/10.1145/2723872.2723884 (cit. on p. 42).

[151] *Occam.* [Online]. Available: https://occam.cs.pitt.edu/ (cit. on p. 42).

[152] M. A. Heroux, "Toward resilient algorithms and applications," Mar. 2014. [Online]. Available: https://arxiv.org/pdf/1402.3809.pdf (cit. on p. 42).

[153] R. Mayer and A. Rauber, "A quantitative study on the re-executability of publicly shared scientific workflows," in *2015 IEEE 11th International Conference on e-Science*, Aug. 2015, pp. 312–321. DOI: 10.1109/eScience.2015.58 (cit. on p. 42).

[154] L. A. Barba, "The hard road to reproducibility," *Science*, vol. 354, no. 6308, pp. 142–142, 2016, ISSN: 0036-8075. DOI: 10.1126/science.354.6308.142. eprint: http://science.sciencemag.org/content/354/6308/142.full.pdf. [Online]. Available: http://science.sciencemag.org/content/354/6308/142 (cit. on p. 42).

[155] A. Rowhani-Farid and A. G. Barnett, "Has open data arrived at the british medical journal (bmj)? an observational study," *BMJ Open*, vol. 6, no. 10, 2016, ISSN: 2044-6055. DOI: 10.1136/bmjopen-2016-011784. eprint: http://bmjopen.bmj.com/content/6/10/e011784.full.pdf. [Online]. Available: http://bmjopen.bmj.com/content/6/10/e011784 (cit. on p. 42).

[156] *Top guidelines*. [Online]. Available: https://cos.io/our-services/top-guidelines/ (cit. on p. 42).

[157] A. H. Baker, D. M. Hammerling, M. N. Levy, H. Xu, J. M. Dennis, B. E. Eaton, J. Edwards, C. Hannay, S. A. Mickelson, R. B. Neale, D. Nychka, J. Shollenberger, J. Tribbia, M. Vertenstein, and D. Williamson, "A new ensemble-based consistency test for the community earth system model (pycect v1.0)," *Geoscientific Model Development*, vol. 8, no. 9, pp. 2829–2840, 2015. DOI: 10.5194/gmd-8-2829-2015. [Online]. Available: https://www.geosci-model-dev.net/8/2829/2015/ (cit. on p. 43).

[158] A. H. Baker, Y. Hu, D. M. Hammerling, Y.-H. Tseng, H. Xu, X. Huang, F. O. Bryan, and G. Yang, "Evaluating statistical consistency in the ocean model component of the community earth system model (pycect v2.0)," *Geoscientific Model Development*, vol. 9, no. 7, pp. 2391–2406, 2016. DOI: 10.5194/gmd-9-2391-2016. [Online]. Available: https://www.geosci-model-dev.net/9/2391/2016/ (cit. on p. 43).

[159] D. J. Milroy, A. H. Baker, D. M. Hammerling, J. M. Dennis, S. A. Mickelson, and E. R. Jessup, "Towards characterizing the variability of statistically consistent community earth system model simulations," *Procedia Computer Science*, vol. 80, pp. 1589–1600, 2016, International Conference on Computational Science 2016, ICCS 2016, 6-8 June 2016, San Diego, California, USA, ISSN: 1877-0509. DOI: http://dx.doi.org/10.1016/j.procs.2016.05.489. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050916309759 (cit. on p. 43).

[160] R. M. Wakimoto, *Dear colleague letter: Reproducibility and robustness of results*, NSF 16-083, May 2016. [Online]. Available: https://www.nsf.gov/pubs/2016/nsf16083/nsf16083.jsp (cit. on p. 43).

[161] F. L. Cook, *Dear colleague letter: Robust and reliable research in the social, behavioral, and economic sciences*, NSF 16-137, Sep. 2016. [Online]. Available: https://www.nsf.gov/pubs/2016/nsf16137/nsf16137.jsp (cit. on p. 43).

[162] J. Kurose, *Dear colleague letter: Encouraging reproducibility in computing and communications research*, NSF 17-022, Oct. 2016. [Online]. Available: https://www.nsf.gov/pubs/2017/nsf17022/nsf17022.jsp (cit. on p. 43).

[163] P. Ivie, C. Zheng, and D. Thain, *A first look at reproducibility and non-determinism in cms software and root data*, 2016. [Online]. Available: http://www3.nd.edu/~dthain/talks/repro-chep-2016.pptx (cit. on p. 44).

[164] M. Walfish and A. J. Blumberg, "Verifying computations without reexecuting them," *Commun. ACM*, vol. 58, no. 2, pp. 74–84, Jan. 2015, ISSN: 0001-0782. DOI: 10.1145/2641562. [Online]. Available: http://doi.acm.org/10.1145/2641562 (cit. on p. 44).

[165] B. Shneiderman, "Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight," *Proceedings of the National Academy of Sciences*, vol. 113, no. 48, pp. 13 538–13 540, 2016. DOI: 10.1073/pnas.1618211113. eprint: http://www.pnas.org/content/113/48/13538.full.pdf. [Online]. Available: http://www.pnas.org/content/113/48/13538.short (cit. on p. 44).

[166] K. Krippendorff, "Computing krippendorff's alpha-reliability," 2011. [Online]. Available: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1043%5C&context=asc%5C_papers (cit. on p. 44).

[167] *Neurofinder*. [Online]. Available: https://neurofinder.codeneuro.org (cit. on p. 44).

[168] M. Broussard, "Big data in practice," *Digital Journalism*, vol. 4, no. 2, pp. 266–279, 2016. DOI: 10.1080/21670811.2015.1074863. eprint: http://dx.doi.org/10.1080/21670811.2015.1074863. [Online]. Available: http://dx.doi.org/10.1080/21670811.2015.1074863 (cit. on p. 44).

[169] *The world factbook*. [Online]. Available: https://www.cia.gov/library/publications/the-world-factbook/appendix/appendix-d.html (cit. on p. 44).

[170] *Cyclomatic complexity*. [Online]. Available: https://en.wikipedia.org/wiki/Cyclomatic%5C_complexity (cit. on p. 45).

[171] L. Hatton, "Power-law distributions of component size in general software systems," *IEEE Transactions on Software Engineering*, vol. 35, no. 4, pp. 566–572, Jul. 2009, ISSN: 0098-5589. DOI: 10.1109/TSE.2008.105 (cit. on p. 45).

[172] ——, "Reexamining the fault density component size connection," *IEEE Software*, vol. 14, no. 2, pp. 89–97, Mar. 1997, ISSN: 0740-7459. DOI: 10.1109/52.582978 (cit. on p. 45).

[173] D. C. Ince, L. Hatton, and J. Graham-Cumming, "The case for open computer programs," *Nature*, vol. 482, no. 7386, pp. 485–488, Feb. 2012, ISSN: 0028-0836. DOI: 10.1038/nature10836. [Online]. Available: http://dx.doi.org/10.1038/nature10836 (cit. on p. 45).

[174] A. C.-A. Sascha Hunold, "Mpi benchmarking revisited: Experimental design and reproducibility," May 2015. eprint: https://arxiv.org/abs/1505.07734 (cit. on p. 45).

[175] K. Bollen, J. T. Cacioppo, R. M. Kaplan, J. A. Krosnick, and J. L. Olds, "Social, behavioral, and economic sciences perspectives on robust and reliable science," May 2015. [Online]. Available: https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf (cit. on p. 45).

[176] B. K. Beaulieu-Jones and C. S. Greene, "Reproducibility of computational workflows is automated using continuous analysis," *Nat Biotech*, vol. 35, no. 4, pp. 342–346, Apr. 2017, Research, ISSN: 1087-0156. [Online]. Available: http://dx.doi.org/10.1038/nbt.3780 (cit. on p. 46).

[177] L. Benedicic, F. A. Cruz, A. Madonna, and K. Mariotti, "Portable, high-performance containers for hpc," Apr. 2017. [Online]. Available: https://arxiv.org/abs/1704.03383 (cit. on p. 46).

[178] A. Rokem, "Reproducibility in human neuroimaging research: A practical example from the analysis of diffusion mri," [Online]. Available: https://www.practicereproducibleresearch.org/case-studies/arokem.html (cit. on p. 46).

[179] S. Hunold, A. Carpen-Amarie, and J. L. Träff, "Reproducible mpi micro-benchmarking isn't as easy as you think," in *Proceedings of the 21st European MPI Users' Group Meeting*, ser. EuroMPI/ASIA '14, Kyoto, Japan: ACM, 2014, 69:69–69:76, ISBN: 978-1-4503-2875-3. DOI: 10.1145/2642769.2642785. [Online]. Available: http://doi.acm.org/10.1145/2642769.2642785 (cit. on p. 46).

[180] A. Carpen-Amarie, S. Hunold, and J. L. Träff, "On the expected and observed communication performance with mpi derived datatypes," in *Proceedings of the 23rd European MPI Users' Group Meeting*, ser. EuroMPI 2016, Edinburgh, United Kingdom: ACM, 2016, pp. 108–120, ISBN: 978-1-4503-4234-6. DOI: 10.1145/2966884.2966905. [Online]. Available: http://doi.acm.org/10.1145/2966884.2966905 (cit. on p. 46).

[181] S. Hunold, A. Carpen-Amarie, F. D. Lübbe, and J. L. Träff, "Automatic verification of self-consistent mpi performance guidelines," in *Proceedings of the 22Nd International Conference on Euro-Par 2016: Parallel Processing - Volume 9833*, New York, NY, USA: Springer-Verlag New York, Inc., 2016, pp. 433–446, ISBN: 978-3-319-43658-6. DOI: 10.1007/978-3-319-43659-3_32. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-43659-3_32 (cit. on p. 46).

[182] *Reprompi benchmark*. [Online]. Available: https://github.com/hunsa/reprompi (cit. on p. 46).

[183] A. Carpen-Amarie, A. Rougier, and F. D. Lübbe, "Stepping stones to reproducible research: A study of current practices in parallel computing," in *Euro-Par 2014: Parallel Processing Workshops: Euro-Par 2014 International Workshops, Porto, Portugal, August 25-26, 2014, Revised Selected Papers, Part I*, L. Lopes, J. Žilinskas, A. Costan, R. G. Cascella, G. Kecskemeti, E. Jeannot, M. Cannataro, L. Ricci, S. Benkner, S. Petit, V. Scarano, J. Gracia, S. Hunold, S. L. Scott, S. Lankes, C. Lengauer, J. Carretero, J. Breitbart, and M. Alexander, Eds. Cham: Springer International Publishing, 2014, pp. 499–510, ISBN: 978-3-319-14325-5. DOI: 10.1007/978-3-319-14325-5_43. [Online]. Available: https://doi.org/10.1007/978-3-319-14325-5%5C_43 (cit. on p. 47).

[184] S. B. Nissen, T. Magidson, K. Gross, and C. T. Bergstrom, "Research: Publication bias and the canonization of false facts," *eLife*, Dec. 2016. DOI: 10.7554/eLife.21451. [Online]. Available: https://elifesciences.org/articles/21451 (cit. on p. 47).

[185] G. Chang, C. B. Roth, C. L. Reyes, O. Pornillos, Y.-J. Chen, and A. P. Chen, "Retraction," *Science*, vol. 314, no. 5807, pp. 1875–1875, 2006, ISSN: 0036-8075. DOI: 10.1126/science.314.5807.1875b. eprint: http://science.sciencemag.org/content/314/5807/1875.2.full.pdf. [Online]. Available: http://science.sciencemag.org/content/314/5807/1875.2 (cit. on p. 47).

[186] G. Miller, "A scientist's nightmare: Software problem leads to five retractions," *Science*, vol. 314, no. 5807, pp. 1856–1857, 2006, ISSN: 0036-8075. DOI: 10.1126/science.314.5807.1856. eprint: http://science.sciencemag.org/content/314/5807/1856.full.pdf. [Online]. Available: http://science.sciencemag.org/content/314/5807/1856 (cit. on p. 47).

[187] V. Stodden and S. Miguez, "Provisioning reproducible computational science," Apr. 2014. [Online]. Available: https://www.xsede.org/documents/659353/703287/xsede14_stodden.pdf (cit. on p. 47).

[188] S. J. Eglen, B. Marwick, Y. O. Halchenko, M. Hanke, S. Sufi, P. Gleeson, R. A. Silver, A. P. Davison, L. Lanyon, M. Abrams, T. Wachtler, D. J. Willshaw, C. Pouzat, and J.-B. Poline, "Toward standard practices for sharing computer code and programs in neuroscience," *Nat Neurosci*, vol. 20, no. 6, pp. 770–773, Jun. 2017, Commentary, ISSN: 1097-6256. [Online]. Available: http://dx.doi.org/10.1038/nn.4550 (cit. on p. 47).

[189] "Extending transparency to code," *Nat Neurosci*, vol. 20, no. 6, pp. 761–761, Jun. 2017, Editorial, ISSN: 1097-6256. [Online]. Available: http://dx.doi.org/10.1038/nn.4579 (cit. on pp. 47, 48).

[190] T. Hothorn and F. Leisch, "Case studies in reproducibility," *Briefings in Bioinformatics*, vol. 12, no. 3, pp. 288–300, 2011. DOI: 10.1093/bib/bbq084. eprint: /oup/backfile/content_public/journal/bib/12/3/10.1093/bib/bbq084/2/bbq084.pdf. [Online]. Available: +%20http://dx.doi.org/10.1093/bib/bbq084 (cit. on p. 48).

[191] C. Trippel, Y. A. Manerkar, D. Lustig, M. Pellauer, and M. Martonosi, "Tricheck: Memory model verification at the trisection of software, hardware, and isa," in *Proceedings of the Twenty-Second International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '17, Xi'an, China: ACM, 2017, pp. 119–133, ISBN: 978-1-4503-4465-4. DOI: 10.1145/3037697.3037719. [Online]. Available: http://doi.acm.org/10.1145/3037697.3037719 (cit. on p. 48).

[192] L. Lamport, "How to make a multiprocessor computer that correctly executes multiprocess programs," *IEEE Transactions on Computers*, vol. C-28, no. 9, pp. 690–691, Sep. 1979, ISSN: 0018-9340. DOI: 10.1109/TC.1979.1675439 (cit. on p. 48).

[193] A. H. M. de Oliveira, D. de Oliveira, and M. Mattoso, "Clouds and reproducibility: A way to go to scientific experiments?" In *Cloud Computing: Principles, Systems and Applications*, N. Antonopoulos and L. Gillam, Eds. Cham: Springer International Publishing, 2017, pp. 127–151, ISBN: 978-3-319-54645-2. DOI: 10.1007/978-3-319-54645-2_5. [Online]. Available: https://doi.org/10.1007/978-3-319-54645-2_5 (cit. on p. 48).

[194] Y. Gil and D. Garijo, "Towards automating data narratives," in *Proceedings of the 22Nd International Conference on Intelligent User Interfaces*, ser. IUI '17, Limassol, Cyprus: ACM, 2017, pp. 565–576, ISBN: 978-1-4503-4348-0. DOI: 10.1145/3025171.3025193. [Online]. Available: http://doi.acm.org/10.1145/3025171.3025193 (cit. on p. 48).

[195] *Datasets for the journal computer physics communications.* [Online]. Available: https://data.mendeley.com/datasets/journals/00104655 (cit. on p. 49).

[196] W.-F. Chiang, G. Gopalakrishnan, Z. Rakamarić, D. H. Ahn, and G. L. Lee, "Determinism and reproducibility in large-scale hpc systems," 2013. [Online]. Available: https://www.semanticscholar.org/paper/Determinism-and-Reproducibility-in-Large-Scale-HPC-Chiang-Gopalakrishnan/9e8ea7d54dc67f672b31b223ed14edc758b0b28d (cit. on p. 49).

[197] *Standing together for reproducibility in large-scale computing*, Dec. 2014. [Online]. Available: https://www.xsede.org/documents/659353/d90df1cb-62b5-47c7-9936-2de11113a40f (cit. on p. 49).

[198] P. Ginsparg, "Next-generation implications of open access," Aug. 2007. [Online]. Available: http://www.ctwatch.org/quarterly/articles/2007/08/next-generation-implications-of-open-access/index.html (cit. on p. 49).

[199] S. A. Iqbal, J. D. Wallach, M. J. Khoury, S. D. Schully, and J. P. A. Ioannidis, "Reproducible research practices and transparency across the biomedical literature," *PLOS Biology*, vol. 14, no. 1, pp. 1–13, Jan. 2016. DOI: 10.1371/journal.pbio.1002333. [Online]. Available: https://doi.org/10.1371/journal.pbio.1002333 (cit. on p. 49).

[200] C. J. Patel, B. Burford, and J. P. A. Ioannidis, "Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations," *Journal of Clinical Epidemiology*, vol. 68, no. 9, pp. 1046–1058, 2017/09/07 XXXX, ISSN: 0895-4356. DOI: 10.1016/j.jclinepi.2015.05.029. [Online]. Available: http://dx.doi.org/10.1016/j.jclinepi.2015.05.029 (cit. on p. 49).

[201] J. P. A. Ioannidis, "Anticipating consequences of sharing raw data and code and of awarding badges for sharing," *Journal of Clinical Epidemiology*, vol. 70, pp. 258–260, 2017/09/07 XXXX, ISSN: 0895-4356. DOI: 10.1016/j.jclinepi.2015.04.015. [Online]. Available: http://dx.doi.org/10.1016/j.jclinepi.2015.04.015 (cit. on p. 49).

[202] A. M. Smith, D. S. Katz, and K. E. a. Niemeyer, "Software citation principles," *PeerJ Computer Science*, vol. 2, e86, Sep. 2016, ISSN: 2376-5992. DOI: 10.7717/peerj-cs.86. [Online]. Available: https://doi.org/10.7717/peerj-cs.86 (cit. on p. 50).

[203] P. Vandewalle, J. Kovacevic, and M. Vetterli, "Reproducible research in signal processing," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 37–47, May 2009, ISSN: 1053-5888. DOI: 10.1109/MSP.2009.932122 (cit. on p. 50).

[204] R. Nuzzo, "How scientists fool themselves – and how they can stop," *Nature*, vol. 526, pp. 182–185, Oct. 2015. DOI: 10.1038/526182a. [Online]. Available: http://dx.doi.org/10.1038/526182a (cit. on p. 50).

[205] R. Gentleman and D. T. Lang, "Statistical analyses and reproducible research," *Journal of Computational and Graphical Statistics*, vol. 16, no. 1, pp. 1–23, 2007. DOI: 10.1198/106186007X178663. eprint: http://dx.doi.org/10.1198/106186007X178663. [Online]. Available: http://dx.doi.org/10.1198/106186007X178663 (cit. on p. 50).

[206] K. Danzer and L. A. Currie, "Guidelines for calibration in analytical chemistry. part i. fundamentals and single component calibration (iupac recommendations 1998)," *Pure Appl. Chem.*, vol. 70, no. 4, pp. 993–1014, 1998. DOI: 10.1351/pac199870040993 (cit. on p. 51).

[207] L. Hutchinson and R. Kirk, "High drug attrition rates—where are we going wrong?" *Nat Rev Clin Oncol*, vol. 8, no. 4, pp. 189–190, Apr. 2011, ISSN: 1759-4774. DOI: 10.1038/nrclinonc.2011.34. [Online]. Available: http://dx.doi.org/10.1038/nrclinonc.2011.34 (cit. on p. 51).

[208] D. Fanelli, "Do pressures to publish increase scientists' bias? an empirical support from us states data," *PLOS ONE*, vol. 5, no. 4, pp. 1–7, Apr. 2010. DOI: `10.1371/journal.pone.0010271`. [Online]. Available: `https://doi.org/10.1371/journal.pone.0010271` (cit. on p. 51).

[209] L. Osherovich, "Hedging against academic risk," *SciBX*, vol. 4, no. 15, Apr. 2011. DOI: `10.1038/scibx.2011.416` (cit. on p. 51).

[210] S. Schroter, N. Black, S. Evans, F. Godlee, L. Osorio, and R. Smith, "What errors do peer reviewers detect, and does training improve their ability to detect them?" *Journal of the Royal Society of Medicine*, vol. 101, no. 10, pp. 507–514, 2008, PMID: 18840867. DOI: `10.1258/jrsm.2008.080062`. eprint: `https://doi.org/10.1258/jrsm.2008.080062`. [Online]. Available: `https://doi.org/10.1258/jrsm.2008.080062` (cit. on p. 52).

[211] B. Nemery, "What happens to the manuscripts that have not been accepted for publication in occupational and environmental medicine?" *Occupational and Environmental Medicine*, vol. 58, no. 9, pp. 604–607, 2001, ISSN: 1351-0711. DOI: `10.1136/oem.58.9.604`. eprint: `http://oem.bmj.com/content/58/9/604.full.pdf`. [Online]. Available: `http://oem.bmj.com/content/58/9/604` (cit. on p. 52).

[212] R. McDonald, H. Cloft, and D. Kallmes, "Fate of submitted manuscripts rejected from the american journal of neuroradiology: Outcomes and commentary," *American Journal of Neuroradiology*, vol. 28, no. 8, pp. 1430–1434, 2007, ISSN: 0195-6108. DOI: `10.3174/ajnr.A0766`. eprint: `http://www.ajnr.org/content/28/8/1430.full.pdf`. [Online]. Available: `http://www.ajnr.org/content/28/8/1430` (cit. on p. 52).

[213] R. K. Harrison, "Phase ii and phase iii failures: 2013-2015," *Nat Rev Drug Discov*, vol. 15, no. 12, pp. 817–818, Dec. 2016, News and Analysis, ISSN: 1474-1776. DOI: `10.1038/nrd.2016.184`. [Online]. Available: `http://dx.doi.org/10.1038/nrd.2016.184` (cit. on p. 52).

[214] T. C. Hales, "Mathematics in the age of the turing machine," 2013. eprint: `https://arxiv.org/pdf/1302.2898.pdf` (cit. on p. 53).

[215] N. Barnes, "Publish your computer code: It is good enough," *Nature*, vol. 467, 2010. DOI: `10.1038/467753a`. [Online]. Available: `http://dx.doi.org/10.1038/467753a` (cit. on p. 53).

Index

Glossary

Categorize Indicates a source I need to categorize. 17, 18, 24, 25, 35–53

Complete Indicates a section I need to complete. 3–6, 8, 10, 11, 13, 16–23, 35, 36

Expand Indicates a bit of information or part of a section which needs expansion to its own section. 19, 20

More Cites Indicates a citation which I have read at least some of which contains more references which are good for reproducibility but which some have not yet made it into the main bibliography. Score this article for the citations.. 27, 30

Needs Citation Indicates a claim or which I need to find a citation. 3, 5, 9, 10, 16–21, 47

Read Indicates a source I need to read. 13, 17, 18, 20, 24–26, 34–53

Read More Indicates a citation which I read parts of but which I may need to read more.. 12, 25, 29, 30, 34, 50