Reproducible Research

Links and info about reproducible research

Links

Other people and labs doing RR efforts

- Reproducible electronic documents: Jon Claerbout and his colleagues at the Stanford Exploration Project initiated (to our knowledge) the discussions about reproducible research.
- Wavelab: David Donoho and his colleagues at the Stanford Statistics Department
  developed Matlab code to reproduce their results on wavelets.
- Reproducible Neurophysiological Data Analysis: a page by Christophe Pouzat on reproducible research in neurophysiology using R and Sweave.
- Sensorscope: the wireless environmental sensing network developed at EPFL. Detailed descriptions of the sensor platform are available for those interested to reproduce the setup. Documented datasets are also available for
  people interested to reuse the data.
- Xin Li's source code collection for reproducible research, with links to code for various image processing algorithms.
- Al Hero's lab applies reproducible research for their publications.
- Andrew Davison at CNRS works on facilitating reproducible simulations using Python.
- MD Anderson Cancer Center: Bioinformatics hosts supplementary material for a number of their publications including code and data.
- StatReport, a description of reproducible statistical reporting at the Department of Biostatistics, Vanderbilt University.
- Jon Wellner's page with links about reproducible research.
- VisionBib.Com contains a very large bibliography of computer vision papers, as well as listings of vision-related code and datasets.
- Neil Lawrence's lab has reproducible documents and software on machine learning.
- Literate Programming: Don Knuth's original work on how to make code human-readable and a lot of related things.
- eScience Institute: eScience institute at university of Washington.
- Bob: a signal-processing and machine learning toolbox originally developed by the Biometrics Group at Idiap, and used for reproducible research.
- ResearchCompendia.org: a site hosting various research compendia, maintained by Victoria Stodden and colleagues at Columbia University.

Journals with RR initiatives

- ACM: text on result and artifact review and badging for ACM papers, including description on reproducibility.
- Annals of Internal Medicine: When a paper is accepted, the authors are asked explicitly whether their paper is reproducible. If yes, links are provided to the study protocol, data, and/or statistical code.
- Biometrical Journal: Authors are strongly encouraged to submit computer code and data sets used to illustrate new methods. These will be published as supporting information on the journal's webpage once the paper was accepted for publication.
- Biostatistics: papers are labeled with an R if they are reproducible, C if code is available online and D if data is available. Data and code are published on the journal's website.
- IEEE Transactions on Signal Processing: in the acceptance e-mail from the editor-in-chief, the authors are encouraged to make their code and data available online.
- The Insight Journal: An online, open access journal in medical imaging that requires code as
  an integral part of the publication. They also allow for online post-publication reviews.
- IPOL: Image Processing On Line, a journal publishing relevant image processing and image analysis algorithms.

Tools

Open Source

- AMRITA: a cross between a document preparation system, a computational engine, and a programming language.
- Article Authoring Add-in for Microsoft Word: a Microsoft Word add-in that enables more metadata to be captured and stored at the authoring stage and enables semantic information to be preserved through the publishing process, which is essential for enabling search and semantic analysis once the articles are archived within information repositories
- CARE: the comprehensive archiver for reproducible execution.
- Cacher: this package provides tools for caching statistical analyses in key-value databases which can subsequently be distributed over the web.
- CDE: a Linux software packaging tool that enables users to easily reproduce computational experiments and deploy prototype software.
- Clawpack: a reproducible research tool in the development of numerical methods for hyperbolic partial differential equations (PDE) by R. J. LeVeque and others.
- coNCePTuaL: A Network Correctness and Performance Testing Language.
- CWEB: a system for literate programming: structured documentation of code to obtain human-readable programs, by D. Knuth and S. Levy.
- DataCite: helping you to find, access and reuse data.
- Knitr: a package designed to be a transparent engine for dynamic report generation with R.

- [Madagascar](#): an open-source software package for multidimensional data analysis and reproducible computational experiments.
- [Noweb](#): a system for literate programming: structured documentation of code to obtain human-readable programs, by N. Ramsey and others.
- [Orcid](#): Open researcher and contributor ID, a community effort to establish an open, independent registry that is adopted and embraced as the industry's de facto standard.
- [Org-mode](#): a tool for keeping notes, maintaining TODO lists, planning projects, and authoring documents with a fast and effective plain-text system. See also [Babel](#) for its ability to have executable source code in a document.
- [RA](#) (ResearchAssistant): a Java library by Daniel Ramage for creating reproducible experiments in Java.
- [RRepository](#): a repository setup for making reproducible research publications available online, based on [EPrints](#).
- [RunMyCode](#): a web platform where you can create a companion site with a paper to allow others to run the corresponding code.
- [SHARE](#): Sharing Hosted Autonomous Research Environments, a method to provide access to a tool that is otherwise cumbersome to install or configure.
- [SQLShare](#): Database-as-a-Service for Researchers.
- [Sumatra](#): a Python-based tool for managing and tracking projects based on numerical simulation or analysis.
- Sweave
  - [The Sweave Homepage](#): by Friedrich Leisch
  - [A Sweave Demo](#): a package to do literate programming and good documentation using the statistical software R, by Charlie Geyer.
  - [Troubleshooting Sweave](#): by John Cook.
- [StatWeave](#): software whereby you can embed statistical code (e.g., SAS, R, Stata, etc.) into a LaTeX or OpenOffice document. A bit like Sweave, but for more languages, developed by Russell V. Lenth.
- [Subversion (SVN) resources](#): see this page from WhoIsHostingThis for a nice overview.
- [TeXmacs](#): an editing platform with special features for scientists, giving a unified and user friendly framework for editing structured documents with different types of content (text, graphics, mathematics, interactive content, etc.).
- [ThePub](#): an alternative setup for making reproducible research publications available online, using Java.
- [Trident](#): A Scientific Workflow Workbench providing a set of tools based on the Windows Workflow Foundation that addresses scientists' need for a flexible, powerful way to analyze large, diverse datasets. It includes graphical tools for creating, running, managing, and sharing workflows.
- [VCR](#): Verifiable Computational Research, a tool to label and reproduce results.
- Version Control Systems
  - [CVS](#): a centralized version control system to preserve code and track changes.
  - [Git](#): a distributed version control system
  - [Mercurial](#): a distributed version control system
  - [Subversion (SVN)](#): a centralized version control system to preserve code and track changes.

- VisTrails: an open-source scientific workflow and provenance management system
  that provides support for data exploration and visualization.
- Zentity: a research output repository platform that provides a suite of building blocks, tools, and services that help you create and maintain an organization's digital library ecosystem.

Commercial

- Inference: a tool for performing reproducible research from within Microsoft Office (Word, Excel) documents, with links to scripts in Matlab, R, etc.

Courses

- Essentials for Data Support, a course on (online) data management by Research Data Netherlands, also including a part on reproducible research.

Blogs

Reproducible Research

- Reproducible Research Ideas: This site's blog about reproducible research.
- RRPlanet Blog: RRPlanet blog about reproducible research.

Related Topics

- The Endeavour: John D. Cook's blog about statistics, programming, and reproducible research.
- EPrints News: latest news from the developers of the EPrints repository software.
- Open Access Archivangelism: Stevan Harnad's blog about Open Access and related issues.
- The Third Bit: Greg Wilson's blog, also containing comments on reproducible research, life in academia, software engineering, etc.
- Victoria Stodden: blog about internet and democracy, open science, intellectual property, etc.
- RetractionWatch: a blog tracking paper retractions.
- Madagascar Development Blog.
- Google newsgroup on reproducible research.