RDA - Reproducibility Interest Group Case Statement

This interest group seeks to advance and enable reproducibility in research based on or producing datasets. Our goals are to provide community-based recommendations and infrastructure solutions, doing so in coordination with the other RDA Working and Interest Groups where appropriate. Our efforts are also intended to produce Working Groups that can pilot these recommendations and solutions, with an eye to building on existing computational infrastructure.

Objectives

Our primary areas of focus regarding reproducibility are:

1) the persistent linking and availability of data and code (via repositories or other mechanisms) used in the generation of published research results, with the publication itself;

2) the development, encouragement, and adoption of meta-data standards for data and code, especially for those linked to publications;

3) the development, encouragement, and adoption of data and code publication, authorship, and citation practices, especially for those linked to publications;

4) the development and adoption of appropriate tools and computational infrastructure that enable: the sharing of research workflows that permit replication of computational scientific findings; the persistent linking of all digital scholarly objects used to generate research findings such as datasets in repositories; and versioning of digital scholarly object to ensure persistent reproducibility.

5) developing a clear research agenda with two co-informed and mutually reinforcing tracks> the first track involves issues such as measuring the reliability of the scientific record; inherent limitations on reproducibility; operationalizing the definition of reproducibility in different research areas; understanding costs and appropriateness in different research settings; and policy design for reproducibility including privacy protection. The second track focuses on computational solutions to irreproducibility, including but not limited to software development (workflow tracking, provenance, and computational environments, as well as tools within traditional analytical software such as statistical packages); machine architecture for reproducibility computational science; cyberinfrastructure for supporting workflow tracking and linking data, code, and results, and enabling the the execution of code, for verification and reuse.

Participation

The Reproducibility Interest Group is open to all RDA members to participate.

Outcomes

The Reproducibility Interest Group will be considered a success if it develops recommendation documents based on consensus amongst its members. In addition, the Interest Group will be considered a success if one or more Working Groups are created to pilot the recommendations.

Mechanism

The Reproducibility Interest Group will utilize capabilities provided by the RDA platform to communicate and collaborate. These include:

Monthly or bimonthly telecons/webex to with a planned agenda to discuss specific issues;
Asynchronous collaboration using online tools such as google docs/groups, wikis and email list servers;
RDA meetings to hold sessions for face to face interactions and to inform other RDA members of its ongoing activities.

All Reproducibility Interest Group documents will be made publicly available.

Timeline

A tentative timeline is given below, subject to change.

Kickoff meeting at 4th RDA Plenary (September 2014):

Discussing focus areas and use cases, identifying overlap with other RDA Interest and Working Groups;
Technology and Infrastructure Lists

By 5th Plenary:

Identifying focus area and science use cases;
Identifying gaps in reproducibility in use cases;
Planning recommendations and infrastructure solutions;
Preliminary Recommendation Document;
Preliminary new Working Group identification.

Potential Group Members:

Bernard Schutz (co-chair)
Victoria Stodden (co-chair)

FIRST NAME	LAST NAME	EMAIL	TITLE
Bernard	Schutz	bernard.schutz@aei.mpg.de	RDA RepIG co-chair; Professor, School of Physics and Astronomy, Cardiff University, UK
Victoria	Stodden	vcs@stodden.net	RDA RepIG co-chair; Associate Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign
Ed	Seidel	eseidel@ncsa.illinois.edu	Director, National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign
Rob	Pennington	rlpennin@illinois.edu	Deputy Director, National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign
Philip	Stark	stark@stat.berkeley.edu	Professor and Chair, Department of Statistics, UC Berkeley
Matt	Turk	matthewturk@gmail.com	Research Scientist, National Center for Supercomputing Applications (NCSA) and Research Assistant Professor of Astronomy at the University of Illinois at Urbana-Champaign
Michael	Hildreth	mhildret@nd.edu	Professor of Physics, Notredame University; Principal Investigator, The DAta and Software Preservation for Open Science (DASPOS)
Paul	Uhlir	PUhlir@nas.edu	Director of the Board on Research Data and Information (BRDI) at the

			U.S. National Academies
George	Alter	altergc@umich.edu	Director, Inter-university Consortium for Political and Social Research (ICPSR); Professor, Department of History; Research Professor, Population Studies Center, University of Michigan.
David	Banks	banks@stat.duke.edu	Professor of the Practice of Statistics, Department of Statistics, Duke University
Dan	Katz	dsk@uchicago.edu	Senior Fellow in the Computation Institute, University of Chicago and Argonne National Laboratory
Laurence	Loewe	loewe@wisc.edu	Assistant Professor, Laboratory of Genetics and Wisconsin Institute for Discovery, UW-Madison
Andy	Terrel	andy.terrel@gmail.com	Research Associate Scientist for the Computational Hydraulics Group at the Institute for Computational Engineering and Science, University of Texas at Austin
Mahadev (Satya)	Satyanarayanan	satya@cs.cmu.edu	Carnegie Group Professor of Computer Science at Carnegie Mellon University
Chris	Mattman	chris.a.mattmann@jpl.nasa.gov	Chief Architect, Instrument Software and Science Data Systems Section, NASA Jet Propulsion Laboratory; Adjunct Associate Professor, Computer

			Science Department, University of Southern California
Jon	Borwein	jon.borwein@gmail.com	Laureate Professor, School of Mathematical and Physical Sciences, University of Newcastle, NSW, Australia
Karl	Broman	kbroman@biostat.wisc.edu	Professor, Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin–Madison
Amy	Brand	a.brand@digital-science.com	VP Academic & Research Relations, VP North America, Digital Science, Macmillan Publishers
John	Wilbanks	wilbanks@gmail.com	Chief Commons Officer at Sage Bionetworks
Mike	Twidale	twidale@illinois.edu	Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign
Jerry	Reiter	jerry@stat.duke.edu	Mrs. Alexander Hehmeyer Professor of Statistical Science, Duke University
Bertram	Ludaescher	ludaesch@gmail.com	Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign
Ian	Foster	foster@anl.gov	Arthur Holly Compton Distinguished Service Professor, Department of Computer Science, University of Chicago; Director, Computation Institute, University of Chicago
Gabrielle	Allen	gdallen@illinois.edu	Associate Director, National Center for

			Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign
Tina	Gunsalus	gunsalus@illinois.edu	Director, National Center for Professional and Research Ethics, Coordinated Science Lab, University of Ilinois at Urbana-Champaign
Steve	Fienberg	fienberg@stat.cmu.edu	Maurice Falk University Professor of Statistics and Social Science in the Department of Statistics, Carnegie Mellon University
Constantine	Gatsonis	gatsonis@stat.brown.edu	Henry Ledyard Goddard University Professor and Chair, Department of Biostatistics, Brown University
John	Wilkin	jpwilkin@illinois.edu	Juanita J. and Robert E. Simpson Dean of Libraries and University Librarian, University of Illinois at Urbana-Champaign
Rafael	Irizarry	rafa@jimmy.harvard.edu	Biostatistics and Computational Biology, Dana Farber Cancer Center; Professor of Biostatistics, Harvard School of Public Health
Cheng-Soon	Ong	Cheng-Soon.Ong@nicta.com.au	Principal researcher at the Machine Learning Research Group, NICTA, Australia
Christine	O'Keefe	Christine.Okeefe@csiro.au	Program Leader for Decision and User Science in CSIRO Computational Informatics, Australia
Neil	Chue Hong	n.chuehong@software.ac.uk	founding Director of the Software Sustainability Institute, UK

Ruth	Heller	ruheller@gmail.com	Department of Statistics and Operations Research, Tel-Aviv University, Tel-Aviv 6997801, Israel
Andrew	Davison	andrew.davison@unic.cnrs-gif.fr	Senior research scientist in the UNIC lab (Unité de Neuroscience, Information et Complexité) of the CNRS, France
Josh	Greenberg	greenberg@sloan.org	Program Officer, Alfred P. Sloan Foundation
Alan	Karr	karr@rti.org	Director of the Center of Excellence for Complex Data Analysis, RTI International
Tanu	Malik	tanum@ci.uchicago.edu	Research Associate, Computation Institute, University of Chicago
Randy	LeVeque	rjl@uw.edu	Professor of Applied Mathematics, University of Washington
Amy	Nurnberger	anurnberger@columbia.edu	Research Data Manager, Center for Digital Research and Scholarship, Columbia University Libraries
Arfon	Smith	arfon@github.com	Github.com
Michael	McLellan	mmclenna@purdue.edu	Senior Research Scientist (HUBzero, nanoHUB), Research Computing, Purdue University
Crosas	Mercè	mcrosas@iq.harvard.edu	Director of Data Science at the Institute for Quantitative Social Science (IQSS), Harvard University
Carole	Goble	carole.goble@manchester.ac.uk	Professor, School of Computer Science, University of Manchester, UK
Christine	Borgman	borgman@gseis.ucla.edu	Presidential Chair and Professor in Information

			Studies, Graduate School of Education and Information Studies, UCLA
Fernando	Perez	fdo.perez@gmail.com	Research Scientist, Henry H. Wheeler Jr. Brain Imaging Center, U.C. Berkeley; and founder of Ipython and Jupyter
John	Ioannidis	jioannid@stanford.edu	C.F. Rehnborg Chair in Disease Prevention at Stanford University, Professor of Medicine, Professor of Health Research and Policy, Director of the Stanford Prevention Research Center (SPRC) at Stanford University School of Medicine, Professor of Statistics (by courtesy) at Stanford University School of Humanities and Sciences, one of the two Directors of the Meta-Research Innovation Center at Stanford (METRICS)
Eric	Eide	eeide@cs.utah.edu	Research Assistant Professor, School of Computing, The University of Utah
Stuart	Buck	SBuck@arnoldfoundation.org	Vice President of Research Integrity, Laura and John Arnold Foundation
Bill	Gropp	wgropp@illinois.edu	Thomas M. Siebel Chair in Computer Science, Computer Science Department, Computer Science Department, University of Illinois at Urbana-Champaign
Steven	Goodman	steve.goodman@stanford.edu	Professor, Medicine -

			General Medical Disciplines, Stanford Medical School; Professor, Health Research & Policy, Stanford University; Co-Director, Meta-research Innovation Center at Stanford (METRICS)
Allen	Renear	renear@illinois.edu	Dean and Professor, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign
Athol	Kemball	akemball@illinois.edu	Associate Professor of Astronomy, Center Affiliate of NCSA, University of Illinois at Urbana-Champaign