

THE RECOMPUTATION MANIFESTO

IAN P. GENT, APRIL 2013
VERSION 1: *REVISION* : 9478

1. *Computational experiments should be recomputable for all time*
2. *Recomputation of recomputable experiments should be very easy*
3. *It should be easier to make experiments recomputable than not to*
4. *Tools and repositories can help recomputation become standard*
5. *The only way to ensure recomputability is to provide virtual machines*
6. *Runtime performance is a secondary issue*

Replication of scientific experiments is critical to the advance of science.¹

Computer Science has the enormous advantage that computational experiments are inherently highly replicable, since computers are extremely good at doing the same thing over and over again. Unfortunately, the discipline of Computer Science has never treated replication seriously. Experiments are not normally replicated, and are not normally replicable in any meaningful way. There is a recent trend towards making source code for experiments available, but this only addresses the openness of methods. It can be close to impossible to make that code run in the first place, and – in the happy event that is even possible – use that code to produce data akin to that reported in a paper. I propose that this must change for the health of the discipline. I propose that the only credible technique to make experiments truly replicable is to provide copies of virtual machines in which the experiments are validated to run on the day the experiment was done. I propose that tools and repositories should be made available to make this happen. I wish to be one of those who makes it happen.

I should say a word about the name “recomputation”. The natural word ‘replication’ would cause confusion given its well-established technical meaning in computing. I have chosen the word ‘recomputation’ to mean the replication of computational experiments. This is adding a specific meaning to an existing word,² but the new meaning adds only a nuance and also serves usefully to distinguish replication of computational experiments from other types of experiment.

1. COMPUTATIONAL EXPERIMENTS SHOULD BE RECOMPUTABLE FOR ALL TIME.

Why recomputable? And why for all time?

¹As well as being critical, it is a timely issue, as every week it seems that a major opinion piece puts forward the importance of replication. This week it is in *Nature* [15].

²‘Recomputation’ is a word older than the USA, reported by the *Oxford English Dictionary* in 1766. It does have some technical meanings in Computer Science, e.g. [16], but none are as widespread as those for ‘replication’.

The necessity for recomputation is simple. An experiment is used to form scientific conclusions, and those conclusions are never final. They may always be subject to question, so experimental work may need to be repeated. I would even argue that the less experiments are replicated, the *more* important it is that they be recomputable. The reason is that flaws in an experiment that is frequently replicated are likely to come to light, but not so if the experiment is rarely replicated. The more rarely an experiment is repeated, the longer the time lapse until it is, and the less likely a new researcher will be able to replicate it easily without the original experiment being deliberately made recomputable. If the original experiment is flawed in some way, misleading results can lie uncorrected in the literature for years [1].

Should experiments be recomputable for all time? Yes. Computer Science is unique in having this ability. Imagine if physicists had the opportunity to look through Galileo’s telescopes at Jupiter’s moons, but couldn’t be bothered to store them. There is no reasonable limit to what is the useful life of an experiment is, so the simplest answer is to have no limit. Moore’s law, while it lasts, makes it technically feasible to store experiments forever. It’s affordable, because of storage costs. Amazon Glacier now costs 12¢ per year per GB. With decreasing storage costs, it’s reasonable to guess that \$1 per GB would be enough to pay for permanent storage. It’s even affordable to rerun all computational experiments in history regularly. If researchers donated 10% of CPU resources for new experiments to recomputing old ones, all 5 year old experiments could be rerun each year.³

2. RECOMPUTATION OF RECOMPUTABLE EXPERIMENTS SHOULD BE VERY EASY

Recomputation should be very easy. Ideally, very easy means clicking a button. This will not always be possible, but we should make it as easy as we possibly can for experimenters to reproduce their own experiments and those of other researchers. A lovely example of making it easy to rerun existing experiments is provided by the website runmycode.org, which experimenters can use to provide a simple interface to test their code on different inputs [17].

An important point about this manifesto is that by recomputation I focus mainly on the exact replication of an experiment. No other choice is available since (as I argue below) we have to duplicate the exact experimental conditions to ensure that recomputation is possible. Drummond argues powerfully that this - which he calls ‘replicability’ - is the poor cousin of ‘reproducibility’, where experiments are reproduced with changes to factors believed to be insignificant [7]. There is no question that key advances in science should be reproducible in Drummond’s richer sense, but neither do I accept that “the impoverished version, replicability, is one not worth having” [7]. To take a non-computational example, it would be wonderful if the original experiments on cold fusion could be replicated exactly. If there was some technical flaw with the experiments, it could be discovered, or if not the effects that led to the apparent conclusion that cold fusion existed could be investigated. Once we have ensured that experiments can be recomputed exactly, we can have the luxury of seeking to ensure that experiments can be reproduced in richer ways.

3. IT SHOULD BE EASIER TO MAKE EXPERIMENTS RECOMPUTABLE THAN NOT TO

How can it be easier to make new experiments recomputable than not to bother? I draw your attention to Jon Claerbout’s words [4]:

³Assuming computing power doubles every 18 months, CPU resources increase by 10 times every 5 years.

“It’s not really for the benefit of other people. Experience shows the principal beneficiary of reproducible research is you the author yourself.”

I’ve been doing major computational experiments for 20 years: some of them I am rather proud of. But whenever I start on a new set of experiments, my heart sinks a little. I expect the pain of getting everything set up, and then I expect the difficulty or impossibility of reproducing my experiments if, say, the paper is rejected and I need to rerun them for a later submission. We should fix this. Tools should be available to make it easier to run experiments, encourage good experimental practice, and simultaneously make them recomputable. An experienced experimenter’s heart should sink no more than an experienced programmer when starting a new program. Just like programmers have comfortable development environments they can use, experimenters should have a range of environments to develop experiments in. And repositories of past experiments - of one’s own and other people’s - should act like libraries to aid rapid development of new experiments.

4. TOOLS AND REPOSITORIES CAN HELP RECOMPUTATION BECOME STANDARD

I draw an analogy with version control tools. Learning and using subversion, git, or mercurial adds complication to one’s life. But the benefits are enormous, and without question it is easier overall to build and maintain code using source code control systems than without. Similarly, repositories such as github or bitbucket greatly enhance development and sharing of code. Once we have the appropriate tools and repositories we in Computer Science will have an enormous advantage over other sciences. Many sciences have massive online repositories, of course leveraging advances in Computer Science to their benefit. We can be unique in having massive online repositories of fully realised experiments, not just the data resulting from experiments. We should research and make available tools and repositories for the recomputation of scientific experiments. Sadly, tools for recomputation are relatively lacking to date, although there are promising signs. For example, in a single PhD thesis, Philip Guo introduced Burrito, a tool for electronic lab notebooks, and CDE, an automatic packaging tool for experiments in linux [11]. Other important recent tools are “verifiable computational results” [8], and “Sumatra” [6]. As well as runmycode.org, other repositories helping to make experiments replicable are myExperiment.org [14], SHARE [10], and the specialist journal “Image Processing Online” (www.ipol.im) in which each article is accompanied by runnable code.

5. THE *only* WAY TO ENSURE RECOMPUTABILITY IS TO PROVIDE VIRTUAL MACHINES

While the other points of the manifesto state how I think the world should be, this point is an empirical statement about how the world is. There are two reasons that virtual machines are - at least for now - the only way to go. The first is the almost universal (I suspect) experience among anyone who has built software and especially tried to rebuild it later. A build of the resulting program (and hence experiment) can fail due to arbitrary changes in the machine being used. This applies even where the machine is physically the same one that was used months ago to run an experiment the first time. An innocent software update, to a package not obviously used by the build software, can cause disaster. If the machine is not the same one or a clone thereof, all bets are off. The *only* way to ensure an experiment is recomputable is to make available a virtual machine identical to one which was tested and worked when the experiment was originally run.

The second reason is that the available computers and operating systems change over time. It may be true that code you make available today can be built with only minor pain by many people on current computers. That is unlikely to be true in 5 years, and hardly credible in 20.

A classic example illustrates both these reasons. SHRDLU is one of the most famous programs in the history of AI [18]. The complete source code is available, but it cannot be run. Although the physical machine and OS it ran on can be emulated, we do not have the exact machine state that enabled the program to run. Amongst other issues, Terry Winograd changed his own copy of Lisp, changes that are now lost [5]. What we need – and should provide for new experiments – is an exact virtual machine in which the experiment worked. There are isolated cases out there of researchers providing virtual machines for recomputation, such as Brown [2, 3]. We need to take this from an isolated case of a keen researcher pushing the envelope, and turn it into the standard way Computer Scientists do business. Howe makes this case much more extensively, giving 13 reasons why virtual machines in the cloud can improve reproducibility [12].

This does not mean that all uploads and downloads to repositories should be of full virtual machines. For ease of use, as well as reducing bandwidth, it's important that tools provide as many ways as possible of allowing experiments to be uploaded and downloaded.

6. RUNTIME PERFORMANCE IS A SECONDARY ISSUE

This manifesto is about allowing recomputation of experiments. If a measure is deterministic, e.g. the result of a fixed computation, or the number of nodes in a non-random search tree, it should be replicated every time the experiment is recomputed. But we cannot guarantee the same results in terms of non-deterministic measures such as cpu time. By using virtual machines, we may obtain different results as are seen on a physical machine, and future results on virtual machines may differ from current ones. Even where serious efforts are made, no guarantee that run-time results are replicable can be obtained. For example, even though max-clique researchers have a standard methodology for cross-referencing cpu times, Prosser has shown that these results cannot be relied on even approximately [13].

We must tackle the easier problems first. Ensuring recomputation of experiments is not easy, but it is critical and at least achievable. Obtaining meaningfully replicable cpu time comparisons requires significant additional research – if it is even possible. But, by allowing recomputation of experiments, we allow researchers to recompute experiments in a variety of environments, discovering if conclusions about run times are universal or contingent on a particular environment. The critical thing is to preserve scientific experiments. It's unarguable that if we can't recompute an experiment at all, we can't preserve run time performance.

THE RECOMPUTATION.ORG MISSION

I am starting <http://recomputation.org> as repository for replicable experiments, and starting to work on the tools to use on the site. Based on my arguments above I have the following plans. The overriding goal is the following mission statement:

If we can compute your experiment now, anyone can recompute it 20 years from now

The site `recomputation.org` is brand new and (as I write) holds zero experiments. But the following are the principles which we will use in developing it to deliver the mission statement.

1. `recomputation.org` will make available virtual machines or equivalent technology to allow exact recomputation of lodged experiments.
2. `recomputation.org` will make its best efforts to ensure that all experiments which it believes to be recomputable will remain recomputable for all time.
3. `recomputation.org` will always be free to those lodging bona-fide scientific experiments and to those obtaining past experiments, provided that
 - (a) all aspects of the experiments are freely available,
 - (b) the experimenter's contributions are open source, and
 - (c) fees are not charged for the related scientific publication.⁴
4. `recomputation.org` will provide its code and tools using an appropriate open source licence, including server-side code.
5. `recomputation.org` will serve as a testbed for scientific research into issues such as experimental techniques and methodologies.
6. `recomputation.org` will seek to fund its mission through research funding, donations, and charging for services such as lodging non-open source or non-freely available experiments.

VOTE RECOMPUTATION

The recomputation manifesto is similar in spirit to the 'Science Code Manifesto' but addresses an orthogonal issue. The Science Code Manifesto, `sciencecodemanifesto.org`, demands that code used in scientific publications should be made available in an open way for the useful lifetime of the publication. But this is neither a necessary condition for recomputation nor sufficient for it. Closed source experiments can be recomputed if the appropriate environment is provided – e.g. a suitable virtual machine containing the necessary binaries but not source. Arguably recomputability is *more* important for closed source than for open source. But it must not be thought that open source is sufficient for recomputability. The *only* guarantee of recomputability is the exact environment being available. I do endorse the Science Code Manifesto, and initial efforts at `http://recomputation.org` will be focussed on open source efforts only. As well as the scientific desirability of open source, it avoids one form of potential licencing problems in recomputation.

A manifesto is a call that people reading it should vote for your point of view. Don't vote with a signature or a petition. Vote by making your computational experiments recomputable. Do it at `http://recomputation.org`, or at your own web site, or at another repository. But make your experiments recomputable.

⁴The point of the caveats is that free service might not be offered to authors who will not make source code available, or to publications such as journals who charge either authors or readers for access. Equally, as argued earlier there is major value in making non-open experiments recomputable too, so if funding is available it may be appropriate to allow these to be freely deposited also.

REFERENCES

- [1] J. Christopher Beck, Patrick Prosser, and Richard J. Wallace. Trying again to fail-first. In Boi Faltings, Adrian Petcu, François Fages, and Francesca Rossi, editors, *CSCLP*, volume 3419 of *Lecture Notes in Computer Science*, pages 41–55. Springer, 2004.
- [2] C. Titus Brown. Our approach to replication in computational science. <http://ivory.idyll.org/blog/replication-i.html>, 2012.
- [3] C Titus Brown, Adina Howe, Qingpeng Zhang, Alexis B Pyrkosz, and Timothy H Brom. A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. 2012.
- [4] Jon Claerbout. Reproducible computational research: A history of hurdles, mostly overcome. <http://sepwww.stanford.edu/sep/jon/reproducible.html>.
- [5] Semaphore Corporation. SHRDLU resurrection. <http://www.semaphorecorp.com/misc/shrdlu.html>, 2011.
- [6] Andrew Davison. Automated tracking of computational experiments using sumatra. In *EuroSciPy: 3rd European meeting on Python in Science*, 2010.
- [7] Chris Drummond. Replicability is not reproducibility: Nor is it good science. *Proceedings of the Twenty-Sixth International Conference on Machine Learning: Workshop on Evaluation Methods for Machine Learning IV*, 2009.
- [8] Matan Gavish and David L. Donoho. Three dream applications of verifiable computational results. *Computing in Science and Engineering*, 14(4):26–31, 2012.
- [9] I.P. Gent, S.A. Grant, E. MacIntyre, P. Prosser, P. Shaw, B.M. Smith, and T. Walsh. How not to do it. *Research Report Series-University of Leeds School of Computer studies LU SCS RR*, 1997.
- [10] Pieter Van Gorp and Paul W. P. J. Grefen. Supporting the internet-based evaluation of research software with cloud infrastructure. *Software and System Modeling*, 11(1):11–28, 2012.
- [11] Philip J. Guo. Software tools to facilitate research programming. Ph.D. dissertation, Department of Computer Science, Stanford University, 2012.
- [12] Bill Howe. Virtual appliances, cloud computing, and reproducible research. *Computing in Science and Engineering*, 14(4):36–41, 2012.
- [13] Patrick Prosser. Exact algorithms for maximum clique: A computational study. *Algorithms*, 5(4):545–587, 2012.
- [14] David De Roure, Carole A. Goble, and Robert Stevens. The design and realisation of the my_{experiment} virtual research environment for social sharing of workflows. *Future Generation Comp. Syst.*, 25(5):561–567, 2009.
- [15] Jonathan F. Russell. If a job is worth doing, it is worth doing twice. *Nature*, 496:7, April 2013.
- [16] Christian Schulte. *Programming Constraint Services: High-Level Programming of Standard and New Constraint Services*, volume 2302 of *Lecture Notes in Computer Science*. Springer, 2002.
- [17] Victoria Stodden, Christophe Hurlin, and Christophe Perignon. Runmycode.org: A novel dissemination and collaboration platform for executing published computational results. In *eScience*, pages 1–8. IEEE Computer Society, 2012.
- [18] Terry Winograd. Understanding natural language. *Cognitive Psychology*, 3(1):1–191, 1972.

ABOUT THE AUTHOR

Ian Gent is Professor of Computer Science at the University of St Andrews, Scotland. His interest in the proper foundations of empirical science in computing date almost 20 years. He has given tutorials on “Empirical Methods in CS and AI” at conferences such as IJCAI 2001. Of his non peer-reviewed papers, his most cited by far is “How Not To Do It” [9], a collection of embarrassing mistakes he and colleagues have made in computational experiments. To show how good we are at not doing things right, we mis-spelt the name of one of the authors!

ACKNOWLEDGEMENTS

I wish to thank many scientists I have discussed these issues with over the years, for example my fellow authors of [9]. I especially thank Patrick Prosser for his tenacious pursuit of replication of past experiments and discussions of his struggles in achieving them, Ewan

(not Ewen) MacIntyre for accepting our mis-spelling of his name, Adam Barker, and Lars Kotthoff. I also thank more recent colleagues, including Edwin Brady, Chris Jefferson, Steve Linton, Ian Miguel, Pete Nightingale, Karen Petrie, Aaron Quigley, and Jonathan Ward.

IAN P. GENT SCHOOL OF COMPUTER SCIENCE, UNIVERSITY OF ST ANDREWS, FIFE, KY16 9SX, SCOTLAND, UK. IAN.GENT@ST-ANDREWS.AC.UK. [HTTP://WWW.CS.ST-ANDREWS.AC.UK/~IPG](http://www.cs.st-andrews.ac.uk/~ipg)