

## **Response to US RFI: [Public Access to Peer-Reviewed Scholarly Publications, Data and Code Resulting From Federally Funded Research](#)**

### **Organization submitting response: [Research Software Alliance](#)**

Authors: [Michelle Barker](#) (Director, ReSA)\*, Daniel S. Katz (University of Illinois Urbana-Champaign), Neil P. Chue Hong (Director, Software Sustainability Institute - University of Edinburgh), Chris Mentzel (Stanford University), Karthik Ram (University of California, Berkeley), Catherine Jones (Science Technology Facilities Council), Andrew Treloar (Australian Research Data Commons) \*Contact person

The Research Software Alliance (ReSA) welcomes this opportunity to inform approaches for ensuring broad public access to the peer-reviewed scholarly publications, data, and code that result from federally-funded scientific research.

This submission focuses on how improving the recognition and value of research software can increase the access to unclassified published research, digital scientific data, and code supported by the US Government. ReSA is the international organization representing the research software community. ReSA's vision is that research software be recognized and valued as a fundamental and vital component of research worldwide.

Research software is essential for research and is being more strongly recognized globally by researchers. The National Science Foundation (NSF) identifies software as “directly responsible for increased scientific productivity and significant enhancement of researchers’ capabilities” (NSF 2017). National and international policy changes are now needed to escalate this recognition and to increase the impact of the software and code in important research and policy areas.

Responses to the topics of interests identified in this Request for Information follow:

### **Topic 1: What current limitations exist to the effective communication of research outputs (publications, data, and code) and how might communications evolve to accelerate public access while advancing the quality of scientific research?**

Effective communication of code (or more generally, software) as a critical research output remains limited. The recency of focus on this is demonstrated by the fact that the 2013 memorandum, *Increasing Access to the Results of Federally Funded Scientific Research*, and 2016 updates to this memorandum (Sheehan 2016) only mention data and publications; the reference to code has only come in the last few years. Yet the NSF made 18,592 awards totaling \$9.6 billion to projects that included software as a topic in their abstracts between 1995-2016 (Carver et al. 2018).

Increased access to software is a worthwhile aim in itself, but also requires a focus on software quality, software sustainability, training and human capital. Existing work already identifies recommendations that funding agencies and research institutions can

directly implement in areas such as this (Akhmerov et al. 2019; Clément-Fontaine et al. 2019). Other work makes recommendations directly to the research community on how to improve practice (Jiménez et al. 2017); National Information Standards Organization (NISO) has a [working group](#) developing standards for reproducibility badges. There are a number of areas where there are opportunities for change:

1. Software should be recognized as a primary output of research, not as secondary to a publication or data. This requires significant cultural change, supported by enabling policy and processes.
2. Software and code need to be funded and supported, alongside the people who create them. Employment practices need to evolve to ensure that people in the research sector with software expertise are appropriately rewarded and have long-term career paths.
3. Mechanisms are needed to evaluate software and code to promote publication, sustainability, and reuse, and the skillsets of those who develop it.
4. Software citation needs to become a standard practice, to ensure that publications give credit to all contributors and that the software can be accessed and reviewed as part of the review of the primary product. Achievement of this will require a significant change to journal/publisher policies and author culture.

ReSA coordinates work across these areas internationally, bringing together existing research software organizations to collaborate on larger strategic goals. This includes driving the development and adoption of community-agreed principles to increase findability and accessibility of research software through application of the FAIR (Findable, Accessible, Interoperable Reusable) principles (Research Data Alliance 2020b) and provision of advice on software practices to COVID-19 researchers through the [Software Subgroup](#) contributing to the *Research Data Alliance COVID-19 Guidelines and Recommendations* (Research Data Alliance 2020a).

The unprecedented impact of COVID-19 brings the need for increased access to, and communication of, all research outputs starkly to the fore. It is clear that the speed of response to COVID-19 depends on the breadth and speed at which research about COVID-19 develops and is shared. A key factor supporting this research is access to research data, software and code, to accelerate results whilst ensuring transparency and reproducibility of research results. And while there are good examples of where research outputs were already openly accessible at the start of this pandemic, there are too many examples where they were not - potentially causing life-threatening delays.

A May 2020 *Science* article calls for open sharing of COVID-19 modelling code so that the results can be replicated and evaluated (Barton et al. 2020). While the research community has been increasing access to key software and code, the Imperial College epidemic simulation model that is being utilized by government decision-makers

was not publicly available until 28 April 2020 (Carmack n.d.), with Microsoft working with the Imperial College team to enable this (Adam 2020).

**Topic 2: What more can Federal agencies do to make tax-payer funded research results, including peer-reviewed author manuscripts, data, and code funded by the Federal Government, freely and publicly accessible in a way that minimizes delay, maximizes access, and enhances usability?**

There are three types of national initiatives that can promote enhanced access to federally funded research outputs:

- National policies and strategies promoting open access to research outputs
- Network and collaborative initiatives aiming to facilitate open access
- Support for research infrastructure, including repositories and portals.

In the US, the majority of these initiatives focus on enhanced access to manuscripts and data. The recognition of the role of research software in research, the recognition of it as a reusable product, and the funding for it, specifically including for the ongoing maintenance of software initially developed with federal funding, have all failed to keep pace with the scale of use of software in research. A significant increase in focus on software and code is needed to accelerate improvements in this area, to increase return on investment on significant government funding.

To enhance accessibility, Federal Government should focus on the following:

1. Requiring publicly-funded software to be made as available as possible.
2. Supporting the public funding of both the development and maintenance of research software
3. Implementing measures to ensure that research institutions recognize the essential need for both software development and software maintenance and support the people who do this:
4. Including both software development and maintenance as measures of research output
5. Supporting the development of core software expertise through support of inclusive software skills and training programs, including facilitation of communities of practice:

The Federal Government should collaborate with a range of other stakeholders with a focus on increasing access to software and code, including international initiatives such as [Software Heritage](#) and the [Software Preservation Network](#) in the preservation of software source code; [The Carpentries](#) in the advanced training of researchers;

disciplinary initiatives, research institutions and collaborative initiatives. Relevant US organizations that work in related areas include the developing [US Research Software Sustainability Institute](#), which focuses on improving sustainability of research software; the [US Research Software Engineer Association](#), which brings together the US community of people writing and contributing to research software; and the [Academic Data Science Alliance](#), an independent national resource network that enables academic data science leaders, practitioners, and educators to connect and exchange ideas, and to advance the uptake of data science best practices in higher education.

### **Topic 3: How would American science leadership and American competitiveness benefit from immediate access to these resources?**

The US has the opportunity to lead internationally in enabling accessibility of the knowledge, information and data generated by federally funded research, by increasing access to software and code. This will support research integrity, productivity, and efficiency.

Access to research software is crucial to maintaining research integrity, to maintain American science leadership's reputation for transparency and reproducibility. It is too often the case that research publications do not appropriately cite software and code, making reproducing this research challenging, if not impossible (Smith et al. 2016).

American competitiveness in the science field can also be enhanced by increased access to software and code, to attain productivity and efficiency benefits that will become possible when extensive re-use of software becomes the norm. Development of software repositories and archives actively supports reproducibility and re-use. NSF investments in repositories include the NSF SI2/CSSI Software Institutes, which includes work by the Science Gateway Community Institute (SGCI), the Molecular Sciences Software Institute (MolSSI), and the Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP) to increase reuse of gateways software infrastructure (Wilkins-Diehr et al. 2018), molecular sciences software (Krylov et al. 2018), and high energy physics software (The HEP Software Foundation et al. 2019), respectively. In light of this emphasis on reuse, it is essential that any new software proposed in future investment proposals be required to explain within the context of existing software why new work is required.

### **References**

- Adam, David. 2020. "Special Report: The Simulations Driving the World's Response to COVID-19." *Nature* 580 (7803): 316–18. doi.org/10.1038/d41586-020-01003-6.
- Akhmerov, Anton, Maria Cruz, Niels Drost, Cees Hof, Tomas Knapen, Mateusz Kuzak, Carlos Martinez-Ortiz, Yasemin Turkyilmaz-van der Velden, and Ben van Werkhoven. 2019. "Raising the Profile of Research Software," August. <https://doi.org/10.5281/ZENODO.3378572>.
- Barton, C. Michael, Marina Alberti, Daniel Ames, Jo-An Atkinson, Jerad Bales, Edmund

- Burke, Min Chen, et al. 2020. "Call for Transparency of COVID-19 Models." Edited by Jennifer Sills. *Science* 368 (6490): 482.2-483. <https://doi.org/10.1126/science.abb8637>.
- Carmack, John. n.d. "The Imperial College Epidemic Simulation Code That I Helped a Little on Is Now Public:' / Twitter." Twitter. Accessed May 6, 2020. [https://twitter.com/id\\_aa\\_carmack/status/1254872368763277313](https://twitter.com/id_aa_carmack/status/1254872368763277313).
- Carver, Jeffrey C., Sandra Gesing, Daniel S. Katz, Karthik Ram, and Nicholas Weber. 2018. "Conceptualization of a US Research Software Sustainability Institute (URSSI)." *Computing in Science & Engineering* 20 (3): 4–9. <https://doi.org/10.1109/MCSE.2018.03221924>.
- Clément-Fontaine, Mélanie, Roberto Di Cosmo, Bastien Guerry, Patrick MOREAU, and François Pellegrini. 2019. "Encouraging a Wider Usage of Software Derived from Research." Research Report. <https://hal.archives-ouvertes.fr/hal-02545142>.
- Jiménez, Rafael C., Mateusz Kuzak, Monther Alhamdoosh, Michelle Barker, Bérénice Batut, Mikael Borg, Salvador Capella-Gutierrez, et al. 2017. "Four Simple Recommendations to Encourage Best Practices in Research Software." *F1000Research* 6 (June): 876. <https://doi.org/10.12688/f1000research.11407.1>.
- Krylov, Anna, Theresa L. Windus, Taylor Barnes, Eliseo Marin-Rimoldi, Jessica A. Nash, Benjamin Pritchard, Daniel G. A. Smith, et al. 2018. "Perspective: Computational Chemistry Software and Its Advancement as Illustrated through Three Grand Challenge Cases for Molecular Science." *The Journal of Chemical Physics* 149 (18): 180901. <https://doi.org/10.1063/1.5052551>.
- NSF. 2017. "Software Infrastructure for Sustained Innovation (SSE, SSI, S2I2): Software Elements, Frameworks and Institute Conceptualizations." 2017. [https://www.nsf.gov/publications/pub\\_summ.jsp?ods\\_key=nsf17526](https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17526).
- Research Data Alliance. 2020a. "RDA COVID-19 Guidelines and Recommendations." RDA. April 23, 2020. <https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-0>.
- . 2020b. "FAIR4RS WG." April 28, 2020. <https://www.rd-alliance.org/groups/fair-4-research-software-fair4rs-wg>.
- Sheehan, Jeremy. 2016. "Increasing Access to the Results of Federally Funded Science." Whitehouse.Gov. February 22, 2016. <https://obamawhitehouse.archives.gov/blog/2016/02/22/increasing-access-results-federally-funded-science>.
- Smith, Arfon M., Daniel S. Katz, Kyle E. Niemeyer, and FORCE11 Software Citation Working Group. 2016. "Software Citation Principles." *PeerJ Computer Science* 2: e86. <https://doi.org/10.7717/peerj-cs.86>.
- The HEP Software Foundation, Johannes Albrecht, Antonio Augusto Alves, Guilherme Amadio, Giuseppe Andronico, Nguyen Anh-Ky, Laurent Aphecetche, et al. 2019. "A Roadmap for HEP Software and Computing R&D for the 2020s." *Computing and Software for Big Science* 3 (1): 7. [doi.org/10.1007/s41781-018-0018-8](https://doi.org/10.1007/s41781-018-0018-8).
- Wilkins-Diehr, Nancy, Michael Zentner, Marlon Pierce, Maytal Dahan, Katherine Lawrence, Linda Hayden, and Nayiri Mullinix. 2018. "The Science Gateways Community Institute at Two Years." In *Proceedings of the Practice and Experience on Advanced Research Computing*, 1–8. Pittsburgh PA USA: ACM. <https://doi.org/10.1145/3219104.3219142>.