

Written Evidence Submitted by the Software Sustainability Institute (RRE0038)

Summary

- Software is a fundamental factor in computational reproducibility, and widely used in research: 69% of researchers in the UK could not undertake their work without specialist software.
- Software has a part to play in research integrity, by making research transparent and reusable. With many studies, research published without the underlying software used to produce the results is unverifiable. Major bodies including the National Academies of Science, OECD and UNESCO have made recommendations around reproducibility and the importance of sharing data and code.
- There are efforts to improve the practice and incentives at the intersection of reproducibility, research integrity and research software. These include the FAIR Principles for Research Software, the Software Citation Principles, and Software Management Plans.
- To enable systemic change to improve reproducibility and research integrity, the quality and transparency of software must be improved. Research software engineers have a key role to play by making the software used in research more robust and reusable, and helping train researchers in the fundamentals of publishing code so others can review and inspect it.
- The new UK Committee on Research Integrity should focus on understanding the entirety of the research system, and ensure that software, data, models and the many varied elements of contemporary research are considered to ensure that innovative solutions to improving research quality are evidence-based and incentivised.

Introduction

1. The Software Sustainability Institute (SSI; www.software.ac.uk) is the UK's national facility for improving research software practice, funded by all seven UKRI research councils. It is a partnership between the universities of Edinburgh, Manchester, Oxford and Southampton, in collaboration with a worldwide community of researchers, research software engineers, infrastructure providers and policymakers. This document represents the position of the Software Sustainability Institute staff, and does not seek to represent the position of our wider stakeholders, including funders, advisory board members and Fellows.

2. The Software Sustainability Institute was established in 2010 and works to improve software practice amongst the users (researchers) and developers (research software engineers) of software used in research. We believe that better software leads to better research.
3. Software is a fundamental factor in reproducibility. Software is used to simulate phenomena, generate, store and clean data, manage experiments, and analyse and visualise results. Every aspect of research relies on software: a 2014 study conducted by the Software Sustainability Institute showed that for 70% of UK researchers “it would not be practical to conduct my work without software”¹. A similar study of US researchers in 2017 revealed a similar reliance on software².
4. Given the dependence on software, many efforts have been made to understand its impact on reproducibility^{3,4,5}. It has become a topic of interest for professional societies⁶ and standards bodies⁷. Software also has a part to play in research integrity, by making research transparent and reusable. With many studies, research published without the underlying software used to produce the results is unverifiable.
5. The Software Sustainability Institute workshop ran a workshop on reproducibility in 2014⁸ which noted that *“a significant (and increasing) amount of research is fundamentally reliant on results generated by software [but] is infrequently subjected to the same scientific rigour as is applied to more conventional experimental apparatus”* and identified that *“the perennial problem of motivating researchers and those who support them to put in the effort to make their work reproducible is key: this requires a culture change, and culture changes take time. However investing in education, systems of credit for software, advocacy, training of researchers in computational techniques and mandating of publishers to require code and data deposition can be seen as ways forward”*.

Recent changes in policy

6. While the challenges of reproducibility and research integrity have been highlighted for some time, there have been significant changes in policy and recommendations in the last few years that provide levers to encourage the required culture change.
7. A 2019 National Academy of Sciences report on “Reproducibility and Replicability in Science”⁹ notes that *“Reproducibility is strongly associated with transparency; a study’s data and code have to be available in order for others to reproduce and confirm results.”*
8. The 2021 revision by the OECD to the “Recommendation of the Council concerning Access to Research Data from Public Funding”¹⁰ notes that *“Access to data and*

¹ 2014 Software in research survey. <https://doi.org/10.5281/zenodo.1183561>

² Surveying the U.S. National Postdoctoral Association Regarding Software Use and Training in Research. <https://doi.org/10.6084/m9.figshare.5328442>

³ Publishing Standards for Computational Science: “Setting the Default to Reproducible” (2013)

⁴ Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. <http://doi.org/10.5334/jors.ay>

⁵ A computational reproducibility study of PLOS ONE articles featuring longitudinal data analyses. <https://doi.org/10.1371/journal.pone.0251194>

⁶ ACM Reproducibility and Replicability EIG. <https://reproducibility.acm.org/>

⁷ NISO RP-31-2021, Reproducibility Badging and Definitions. <https://doi.org/10.3789/niso-rp-31-2021>

⁸ Software in reproducible research: advice and best practice collected from experiences at the collaborations workshop. <https://doi.org/10.1145/2618137.2618140>

⁹ Reproducibility and Replicability in Science. <https://doi.org/10.17226/25303>

other research-relevant digital objects enhances the reproducibility of scientific results” and recommends OECD members “Take steps to make research data and other research-relevant digital objects from public funding understandable and re-usable in the long term, including through the provision of high quality human-readable, machine-actionable, and open metadata and adequately maintained and supported bespoke algorithms, code, software, and workflows essential for re-use of data as free and open source.”

9. The draft text of the UNESCO Recommendation on Open Science¹¹, to be ratified later in 2021, considers *“that more open, transparent, collaborative and inclusive scientific practices, coupled with more accessible and verifiable scientific knowledge subject to scrutiny and critique, is a more efficient enterprise that improves the quality, reproducibility and impact of science, and thereby the reliability of the evidence needed for robust decision-making and policy and increased trust in science”* and recommends *“Promoting open science from the outset of the research process and extending the principles of openness in all stages of the scientific process to improve quality and reproducibility”*.

Recent advances in practice

10. The Software Sustainability Institute has been involved in a number of international efforts to improve the practices and incentives around software, reproducibility and research integrity. These are integral to lowering the barriers to making research more reproducible, and rewarding those who follow good practice.
11. The “FAIR Principles for Research Software (FAIR4RS)”¹² improve the practice of research and scholarship by making software more findable, accessible, interoperable and reusable and encourages researchers to make their software more transparent. This builds on “The FAIR Guiding Principles for scientific data management and stewardship”¹³ which notes that *“all scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability”*.

¹⁰ Council Recommendation on Access to Research Data from Public Funding. OECD/LEGAL/0347.

¹¹ Draft text of the UNESCO Recommendation on Open Science. CL/4363.

¹² FAIR Principles for Research Software. <https://doi.org/10.15497/RDA00065>

¹³ The FAIR Guiding Principles for scientific data management and stewardship. <https://doi.org/10.1038/sdata.2016.18>

12. The “Software Citation Principles”¹⁴ note that *“citation of specific software used is necessary for reproducibility, although not sufficient”* and *“adherence to the software citation principles enables better peer review through improved reproducibility”*. Recently, through the work of the FORCE11 Software Citation Implementation Working Group, software citation has been supported by major infrastructures such as GitHub¹⁵ and Zenodo, and major academic publishers¹⁶. This increases the credit for publishing the software associated with publications.
13. Software Management Plans¹⁷, like Data Management Plans, are a tool for helping researchers reflect on the outputs they will be producing, and ensuring that these are properly managed, licensed, curated and published. This improves reproducibility and research integrity by encouraging researchers to consider what is required before the research is carried out, similar to the use of registered reports in other fields.
14. The “NISO Recommended Practice on Reproducibility Badging and Definitions”¹⁸ is *“an effort to develop common recognition practices, vocabulary, and iconography used to facilitate the sharing of data and methods”* and offers *“a set of unified recommended practices regarding reproducibility badging”* to *“provide guidance on badging standards to avoid proliferation and conflicts”*. This should improve the uptake of badging to identify where authors have taken steps towards making their work more reproducible in a way that is easily understood by both authors and readers across publications and publishers.

Supporting change

15. Improving reproducibility and research integrity through better software practice will require systemic support. Researchers are - rightly - focused on the advancement of knowledge. They require other professional roles to support them, and training and guidance to allow them to understand and adopt good practice.
16. As all research relies on software and data, researchers today require a new set of foundational skills to supplement existing basic training in topics such as statistics or ethics. The Carpentries¹⁹ is an open, global initiative to teach foundational coding and data science skills to researchers worldwide that has run 2,700 workshops in 71 countries and trained over 2,800 volunteer instructors to deliver courses to 66,000 novice learners. The Software Sustainability Institute has supported the uptake of Carpentries courses in the UK - delivering over 420 workshops and creating 320 new instructors - that has improved the ability of thousands of researchers to make their research more reproducible.

¹⁴ Software citation principles. <https://doi.org/10.7717/peerj-cs.86>

¹⁵ <https://docs.github.com/en/repositories/managing-your-repositorys-settings-and-features/customizing-your-repository/about-citation-files>

¹⁶ Software citation policies index. <https://www.chorusaccess.org/resources/software-citation-policies-index/>

¹⁷ Writing and using a software management plan.

<https://www.software.ac.uk/resources/guides/software-management-plans>

¹⁸ NISO RP-31-2021, Reproducibility Badging and Definitions. <https://doi.org/10.3789/niso-rp-31-2021>

¹⁹ The Carpentries. <https://carpentries.org/>

17. Through the Software Sustainability Institute's Fellowship programme, we have supported the development of The Turing Way²⁰ an open, collaborative and community-driven "*handbook to reproducible, ethical and collaborative data science*". This has been used by institutions across the UK and internationally to instil grassroots culture change and has spawned book dashes and community co-working calls.
18. The Software Sustainability Institute has also supported CodeCheck²¹ which tackles one of the main challenges to verifying reproducibility by supporting codecheckers with a workflow, guidelines and tools to evaluate computer programs underlying scientific papers, enabling the award a "certificate of executable computation". This approach was used to independently assess Prof. Neil Ferguson's CovidSim models²².
19. One of the main changes in research in the last decade²³ is the recognition of Research Software Engineers as a distinct role in the research process. From a community workshop organised by the Software Sustainability Institute in 2012, this has led to the formation of a professional society, the Society of Research Software Engineering²⁴, an international conference, and the establishment of RSE groups and enhanced career paths at universities. This has helped improve the computational reproducibility of research by increasing software skills across everyone in research and promoting collaboration between researchers and software experts, leading to highly quality, more transparent and more reusable software (and research).

The role of funders

20. Funders have a key role to play in improving the quality of research, because of the incentives created by the design and delivery of funding schemes, and the impact this has on the behaviour of researchers. A key lever, as recognised by the OECD and UNESCO, is requiring sharing of data, code and other materials as appropriate, but this is reliant in incentivising this practice and having the resources to enforce it.
21. Funders must ensure that institutions provide and encourage training in research integrity, open research, and proper computational and data science practices. This should be available at all career stages, and career development plans should become a mandatory part of all funding proposals.
22. Funders could improve the reproducibility of research by directly encouraging the publishing of software and data openly (where allowed). Proper application of machine-actionable data management plans and software management plans are an essential tool to enable this, but it will also require investment in the relevant infrastructure (e.g., digital repositories for software, data, models and electronic notebooks) and training to ensure that published outputs are of sufficient quality and well described.

²⁰ The Turing Way. <https://doi.org/10.5281/zenodo.3233853>

²¹ CodeCheck. <https://codecheck.org.uk/>

²² CODECHECK certificate 2020-010. <https://doi.org/10.5281/zenodo.3865490>

²³ Research Software Engineers: State of the Nation Report 2017.

<https://doi.org/10.5281/zenodo.495360>

²⁴ Society of Research Software Engineering. <https://society-rse.org/>

Conclusions

23. Despite the fears over a lack of reproducibility of research, there are many signs that the research community is willing to address the issues. However much of what is being done is still reactive and placing the burden on individuals rather than assessing where changes to the systems could bring about greater results. The new UK Committee on Research Integrity should therefore focus on understanding the complexities of the research ecosystem, identifying areas for improvement (e.g., adoption of open research practices, promotion criteria, moving away from “hero science” to collaborative teams), and advocating for more research to understand the impact of novel approaches and innovations, enabling an evidenced-based approach to improving reproducibility and research integrity.
24. Fundamentally, reproducibility cannot be improved without increased quality and transparency of software and research integrity always risks compromise if software is not open to scrutiny. To address this, the UK must ensure that all researchers are trained in the basic digital competencies to facilitate modern research methods and are supported by equitable access to specialist competencies (such as research software engineers, data scientists, modellers and research data managers) and well-managed digital infrastructures (such as digital repositories, catalogues, computational platforms, and tools).

(September 2021)