Australian Research Data Commons

A National Agenda for Research Software

By *The Australian Research Data Commons*

March 2022

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

The ARDC is enabled by NCRIS

DOI: 10.5281/zenodo.6378082

Suggested citation:

ARDC Ltd. (2022). A National Agenda for Research Software. Viewed online at: https://doi.org/10.5281/zenodo.6378082

CONTENTS

Cover Image — Siarhei - 338804802 / stock.adobe.com

ARDC is enabled by NCRIS

OUR VISION: RESEARCH SOFTWARE IS RECOGNISED AS A FIRST-CLASS OUTPUT OF RESEARCH

Research software is an essential building block of modern research. The generation, handling and analysis of data almost always involves software. The 2014 UK Research Software Survey found that more than 90% of researchers acknowledged software as being important for their own research, and about 70% of researchers said that their research would not be possible without software. Producing research software is a very common part of research. Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2) included evidence that research software is made across all fields of research, with an average of 33% of research resulting in new code.

Despite this, software is an often invisible part of research, produced quickly within a funding window, often struggling to be maintained beyond that.

> *Without data it's difficult to validate results. But without code, we waste the opportunity to advance science.*
>
> — *Neil Chue Hong, Director, UK Software Sustainability Institute*

Research software is a highly specialised output which has had great impact on society, the economy and industry. It is an output with the same, or even greater potential for impact as journal papers and research data, but it is not treated the same. Recognition of equal treatment is the critical change called for in this agenda.

The vision for the National Agenda for Research Software: Research software is recognised as a first-class output of research

The vision for the National Agenda for Research Software:

RESEARCH SOFTWARE IS RECOGNISED AS A FIRST-CLASS OUTPUT OF RESEARCH.

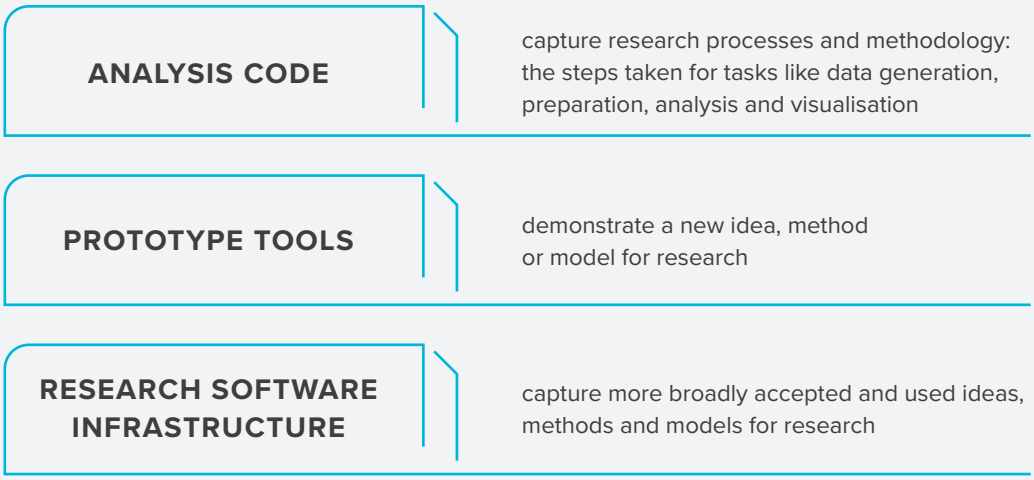
Together we can make this vision a reality through concerted, coordinated action to see, shape and sustain research software.

> *[Research software] is fundamental to research, and it should be treated as a first-class research output to be maintained, assessed and cited, and on an equal footing to research articles and data.*
>
> — *Research Software Current State Assessment, Digital Research Alliance, Canada 2021*

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

WHAT IS RESEARCH SOFTWARE?

Software can come in many forms, including scripts, code, notebooks, computational workflows, libraries, modules, frameworks, utilities and applications. Here we focus instead on what they get built for.



There is a relationship between these 3 kinds of software – software at the top of the table builds on the existence and stability of the software lower down. Conversely the existence and availability of software at the top builds the case for the value of the software built upon.

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

WHO MAKES RESEARCH SOFTWARE FIRST-CLASS?

Recognising research software as first-class will require action across the board. The following roles need to be considered:

CREATORS
develop analysis code

AUTHORS
develop prototype tools

MAINTAINERS
develop and maintain enduring research software infrastructure

USERS
use research software to create outputs and outcomes

SUPPORTERS
assist with the tracking, creation, handling, maintenance, execution, storage and preservation of research software through direct assistance, training or the production of guidance materials

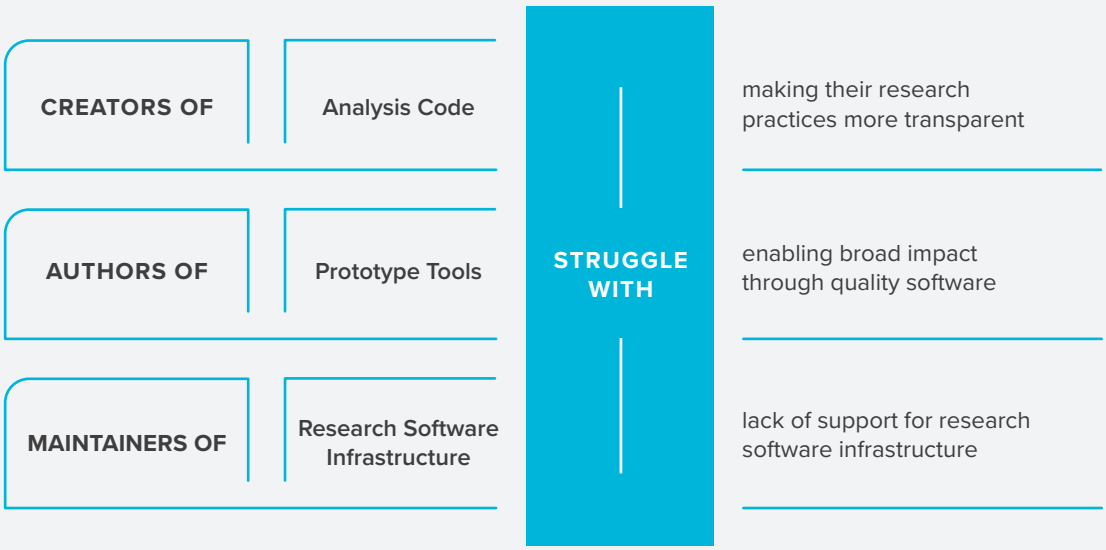
INFRASTRUCTURE PROVIDERS
make or provide the systems that capture or enable the tracking, creation, handling, maintenance, execution, storage and preservation of research software

POLICY MAKERS
formalise norms established by the actions of the people in roles above across communities (research discipline or support groups), organisations (research office, open source program or policy office), infrastructure provision (publishers of research outputs) and funding

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

WHAT ARE THE CHALLENGES?

Recognition of research software occurs under different conditions that give us 3 challenges that must be addressed.



For this agenda, we consider these challenges separately, proposing how to approach them and what the expected outcomes will be. We will see that addressing each challenge supports the work of subsequent challenges.

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

CHALLENGE I:
THE SHIFT TO MORE TRANSPARENT PRACTICES

> *"Thank you for your interest in our paper. For the [redacted] calculations I used my own code, and there is no public version of this code, which could be downloaded. Since this code is not very user-friendly and is under constant development I prefer not to share this code."*
>
> — *anonymised author (and typical) response to a request for code for an article in a journal with a code availability policy*

Everyone who works in research is increasingly hearing the call to work in a more transparent fashion. But the embrace of transparency is at different stages amongst stakeholders, making adoption challenging.

> *"Researchers and institutions are **responsible for fostering transparency** in research."*
>
> — *Publication and dissemination of research: a guide supporting the Australian Code for the Responsible Conduct of Research*

Of importance to research software, transparency includes being open about the methods used in analysis. Capturing methodology means creating a link between the use of software, and the output or outcome it enabled. It can be captured in 1 or more of 3 ways:

- software citation
- as code itself
- provenance metadata.

TRANSPARENCY in declaring interests and reporting research methodology, data and findings

- Share and communicate research methodology, data and findings openly, responsibly and accurately.
- Disclose and manage conflicts of interest.

Principle 3, The Australian Code for the Responsible Conduct of Research, 2018

The practice of software citation is not widespread or consistent. Its use is currently most common amongst those researcher communities that care the most about recognising the value that software represents. Citability is not consistently enabled by creators/authors/maintainers of research software.

Amongst our 3 types of software, the shift to visible analysis code is the most challenging. Analysis code captures steps in analysis and what decisions and transformations have occurred with the data. It is frequently not made available by researchers, let alone shared or published.

Platforms or services often provide tools for transformation, analysis and/or visualisation. Decisions and transformations of data are made when preparing reference data assets. Provenance metadata (possibly in the form of a workflow) capturing software as the means of transformation or analysis is rarely made available by providers of this infrastructure.

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

CHALLENGE I: THE SHIFT TO MORE TRANSPARENT PRACTICES

The underlying systems and processes are not in place to enable the shift. Research software as an output is not usually anticipated, tracked, supported and handled by the institutions and infrastructure providers where research software creators do their work, despite policy makers calling on them to do so.

Approach: See Research Software

Our goal is to see research software in order to recognise it as a first-class output.

We suggest that widespread action is needed to make research software visible.

We will see research software when it is:

- shared, published, or otherwise made available upon creation
- cited or identified in reuse
- captured in data provenance or workflows.

This will require action from a broad range of stakeholders:

- users must cite or otherwise acknowledge their use of research software
- creators/authors/maintainers must make their code available and clarify preferred citation
- supporters and infrastructure providers must enable tracking of outputs, citation, code availability and provenance
- policy makers must formalise emerging norms around visibility
- communities of the above must form to address these changes.

> Develop strategies, including road-maps, funding plans, and business models, to ensure sustainable infrastructures for research data and other research-relevant digital objects from public funding, including data and software repositories and services
>
> *OECD Recommendation concerning Access to Research Data from Public Funding, (revised 2021)*

Outcome: Visible Research Software

Visible research software has the potential to bring broadscale benefit to society, the economy and the environment.

Increased availability of analysis code benefits everyone through more transparent research. Citation, the availability of analysis code, workflows and formal records of software use in data provenance enable a connection between the use of research software and the impact that their work has and enables. This means measures of impact can be established. Overall, research is more accessible and verifiable, increasing trust in research processes and outcomes.

Understanding the breadth and scale of visible research software production enables us to begin to understand the scale of work to be done for our remaining goals.

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

CHALLENGE II:
ENABLING BROAD IMPACT THROUGH QUALITY SOFTWARE

Authors demonstrate new ideas, methods and models arising from research as prototype tools. This kind of software emerges from fixed term research funding, and is often hidden within a larger research project. And then it is hard to find subsequent funding once a prototype has been produced.

In producing this type of research software, the authors must balance applying best practice software development and engineering against producing a functional proof-of-concept within a narrow funding window. In some areas, the problem is compounded by incentives geared towards creating entirely new outputs over contributing additional features to existing ones, leading to duplicated efforts.

Research project leads and/or authors often struggle to connect with either the skills or the experience in producing the software. Authors of this kind of research software are usually highly proficient in their research area (possibly to PhD level), but typically without qualifications or formal training in software development or engineering. This combination of deep research domain knowledge and software development is rare and highly sought after, but not necessarily valued by the systems within which it might emerge.

Approach: Shape Research Software

Our goal is to shape better research software in order to recognise it as a first-class output.

We suggest that widespread action is needed to make research software that is fit-for-purpose.

> *"High quality research software is important for excellence in research. It has become a central component of scientific work as rarely any research is conducted nowadays that does not rely on software."*
>
> — *G6 statement on Open Science, 2021 (Europe)*

We will shape research software when its authoring:

- anticipates broadest appropriate use
- employs best practices for quality software
- where possible, extends upon existing software rather than recreating it

Note that "best practice" in the production of analysis code, prototype tools and research software infrastructure means quite different things.

This clearly depends on software that is more visible. Seeing and valuing research software enables us to make software better.

> We shouldn't be embarrassed by publishing code which is imperfect, nor should other people embarrass us
>
> *The Research Software Impact Manifesto, UK SSI, 2016*

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

This will require action from a broad range of stakeholders:

- users must give feedback on the software they value
- authors/maintainers must gather and act on feedback and learn new skills and/or gain access to skilled supporters
- supporters must teach, guide or provide the relevant skills authors need to make better software
- infrastructure providers must establish and provide the systems and services that enable authors
- policy makers must formalise emerging norms for cutting-edge software outputs that are good quality and fit-for-purpose
- communities of the above must form to address these changes

Outcome: Broadest Appropriate Reuse

Better software means better research. Access to better research software through a culture of building upon existing efforts where appropriate, and developing better and more broadly used prototype tools otherwise will bring broad benefits to society, the economy and the environment.

Research software is a uniquely actionable form of knowledge representation. Unlocking broader audiences for use of this knowledge and ensuring that it is accurate and robust is the best possible base to build towards addressing our final goal: sustainability.

> Better Software, Better Research
>
> *UK Software Sustainability Institute motto*

ARDC is enabled by NCRIS

CHALLENGE III: LACK OF SUPPORT FOR RESEARCH SOFTWARE INFRASTRUCTURE

The maintainers of research software infrastructure enable access to more broadly accepted research methods and models. These maintainers are more likely to have qualifications or extensive experience in software development or engineering, possibly more so than the area or areas of research they are working in.

This work is often conducted on a volunteer basis, or alongside other concerns. A systematic approach to the sustainability of this endeavour simply doesn't exist. There is no bridge between the production of prototype tools and research software infrastructure. Instead this kind of software limps along with success and failure in funding rounds.

In this environment, maintainers struggle to carve out meaningful and sustained careers producing and maintaining research software. Worse yet, this kind of software is predominantly built upon the efforts of a single individual, or to a lesser degree by a small team. If a maintainer leaves the project, the initiative will very likely fail. If this happens, it can leave whole swathes of researchers without the basic tools they need to do their jobs.

Finally, while many researchers agree that long lived research software infrastructure is critical, it is often unclear who holds the responsibility to keep it maintained.

Approach: Sustain Research Software

Our goal is to sustain research software. The ultimate recognition of software as a first-class output would be to provide the support needed to keep it there.

We suggest that widespread action is needed to identify pathways to maintenance of research software and the roles that make this happen.

We sustain research software when it:

- is complemented by roles which are themselves stable, inclusive, supported and valued
- has viable pathways to maintenance.

> *"to have software that people use, is to maintain an act of community"*
>
> — *Dan Simpson, core contributor to Stan, an open source statistical modelling and computation platform*

> These essential staff need reward systems and career pathways that simultaneously support open science practice and their professional development.
>
> *2021 National Research Infrastructure Roadmap Exposure Draft*

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

This will require action from a broad range of stakeholders:

- users must give feedback on the research software they value
- maintainers must work with communities that value their software
- maintainers must have access to the skills and/or support for software engineering
- supporters must teach, guide or provide the skills maintainers need to maintain software and to work with the communities that value their software
- policy makers must formalise the means and mechanisms for maintaining valued research software as pathways emerge
- communities of the above must connect and share knowledge and establish norms and governance.

Outcome: Research Software Sustainability

Sustained research software is a solid base for prototype tools and analysis code to build upon, and for use in broad areas of research. Of all the forms of software discussed, sustained research software is the most likely to be broadly used and benefit society, the economy and the environment.

WHERE TO FROM HERE?

This agenda requires action from a broad range of stakeholders, coming together to work on a common vision. If we work together, we believe that this is a change that can happen over a decade. But with separate, more achievable goals as part of that vision, we believe we will begin to see the change far sooner than that.

For each challenge, we have described goals to see, shape and sustain research software and against each goal, the actions needed by roles that different people perform. If you care about this change, but you're uncertain about where you fit in with those actions, or you would like to reach out to people in similar roles doing similar activities at different organisations, please get in contact with us.

If you'd like to discuss which parts of your organisation need to come together, and what evidence and activity outside your organisation can support you to build your case, please get in contact with us. You may also like to connect with relevant communities, many of which we support or co-lead.

At the ARDC, we're open to partnerships. If you are interested in partnering in action for recognition of research software as a first-class output of research, please get in contact, and watch for opportunities to engage as they arise by signing up to our newsletter.

As part of our commitment to address this agenda, the ARDC is running a rolling series of activities under the research software program. These activities are grouped under 3 projects to see, shape, and sustain research software. We aim to create complementary areas of national infrastructure, guidance, community and advocacy across these 3 projects. Details of concluded and future activities are available via the ARDC website. The evidence, guidance and activities are there to help you make the change.

> Without data it's difficult to validate results. But without code, we waste the opportunity to advance science.
>
> *Neil Chue Hong, Director, UK Software Sustainability Institute*

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

FEEDBACK

We welcome your feedback on this document. Please email *contact@ardc.edu.au* with any comments or questions.

ABOUT THE AUSTRALIAN RESEARCH DATA COMMONS

The Australian Research Data Commons (ARDC) enables the Australian research community and industry access to nationally significant, data intensive digital research infrastructure, platforms, skills and collections of high quality data.

The ARDC is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

NCRIS National Research Infrastructure for Australia An Australian Government Initiative
ARDC is enabled by NCRIS

Australian Research Data Commons

CONTACT

ardc.edu.au
+61 3 9902 0585
contact@ardc.edu.au

FOLLOW

@ardc_au
australian-research-data-commons
subscribe to our newsletter

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

The ARDC is enabled by NCRIS