



Software
Sustainability
Institute

Implementation of policies on research software

Neil Chue Hong (@npch)

Slides: <https://doi.org/10.6084/m9.figshare.21347697>



Supported by:



Research software supports open research



www.software.ac.uk

Research software policy needs to support open research, too.

Implementation of research software policies must be seen as part of the “mosaic” of different research policies:

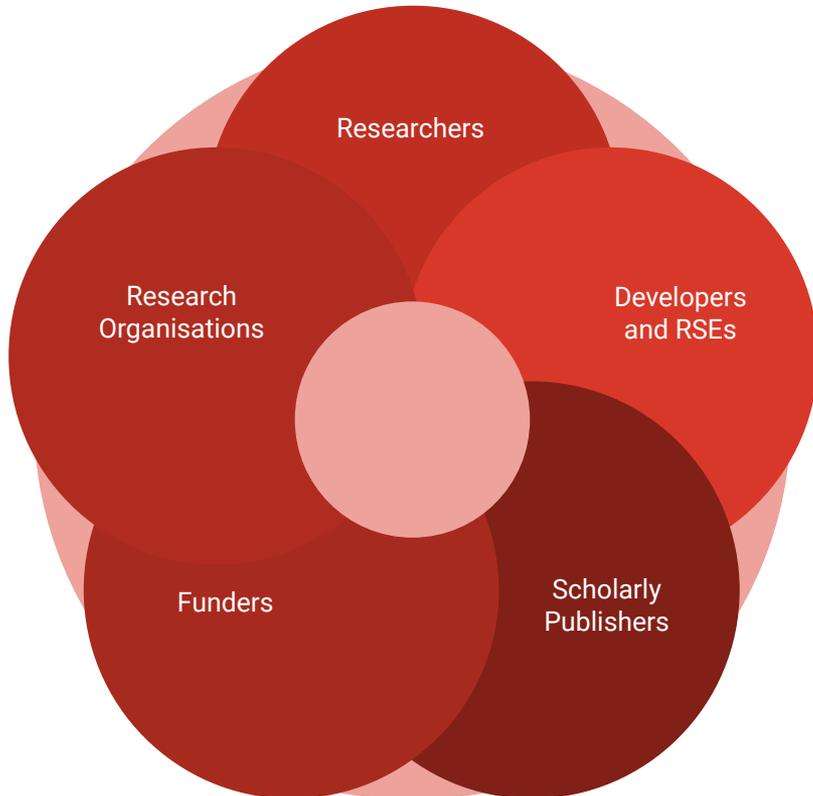
- Open Access
- Research Outcomes / Outputs
- Research Data Management
- Research Ethics and Integrity
- Diversity, Equity and Inclusion in research
- IP and licensing

Making it easier to implement and adopt these policies will help their effectiveness

Stakeholders in research software policy



www.software.ac.uk



The primary stakeholders are:

- Researchers
- Developers / RSEs
- Research Organisations
- Funders
- Scholarly Publishers

Secondary stakeholders include:

- Other users (e.g. industry and public sector); Other policymakers (e.g. government)

Levels of research software policy



www.software.ac.uk

Policy is a deliberate system of guidelines to guide decisions and achieve rational outcomes. A policy is a *statement of intent* and is implemented as a procedure or protocol.

Often linked to governance or a community, e.g.:

- Funder policy on what software licenses grantees are required to use
- University guidelines on the role of software deposit in responsible and reproducible research
- Community software guidelines for research software projects
- Manifestos for environmentally responsible software development

Software is a part of many other research policies



www.software.ac.uk

Many policies affecting research software are not specifically research software policies, e.g.:

- Open Research / Open Science: is predicated on the sharing of code
- Data (or Research Output) Management Plans: will often include software

Others software specific policies may be adapted/derived from other policies and guidelines, often research data, e.g.:

- FAIR Principles for Research Software
- Software Citation Principles

Is one approach better than the other for implementation and adoption?

Policies → Procedures & Protocols



www.software.ac.uk

Open Access policies can be adhered to in different ways e.g. Green OA, Gold OA, etc.

Implementation of research software policies must take into account how they will be implemented.

There is often a gap between policy, procedure, and practice.

This can be seen most clearly when looking at software and data sharing policies.

Effects of changes to Science editorial policy



www.software.ac.uk

In 2011 [Science changed its editorial policies](#): “We require that all computer code used for modeling and/or data analysis that is not commercially available be deposited in a publicly accessible repository upon publication.”

Table 1. Responses to emailed requests (n = 180)

Type of response	Count	Percent, %
Did not share data or code:		
Contact another person	20	11
Asked for reasons	20	11
Refusal to share	12	7
Directed back to supplement	6	3
Unfulfilled promise to follow up	5	3
Impossible to share	3	2
Shared data and code	65	36
Email bounced	3	2
No response	46	26

“Normally we do not provide this kind of information to people we do not know. It might be that you want to check the data analysis, and that might be of some use to us, but only if you publish your findings while properly referring to us.”

“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.”

“I have to say that this is a very unusual request without any explanation! Please ask your supervisor to send me an email with a detailed, and I mean detailed, explanation.”

“When you approach a PI for the source codes and raw data, you better explain who you are, whom you work for, why you need the data and what you are going to do with it.”

Stodden, Seiler, Ma. (2016). An empirical analysis of journal policy effectiveness for computational reproducibility
<https://doi.org/10.1073/pnas.1708290115>

Software sharing for infectious disease models



www.software.ac.uk

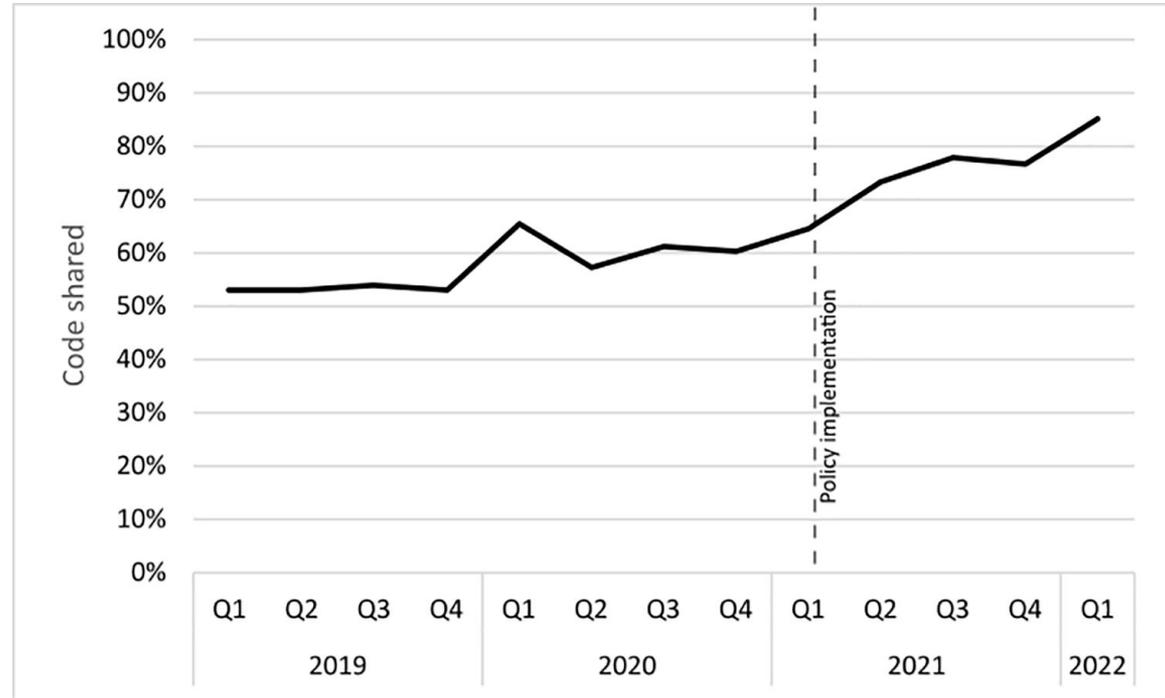
About $\frac{1}{5}$ of infectious disease models published in PubMedCentral shared the associated code, though this has increased due to COVID-19 research.

N = 1338	Code sharing	Data sharing	Registration	COI	Funding
	N (%)	N (%)	N (%)	N (%)	N (%)
Overall	288 (21.5)	332 (24.8)	6 (0.4)	1197 (89.5)	1109 (82.9)
2019	38 (17.6)	59 (27.3)	3 (1.4)	197 (91.2)	202 (93.5)
2021	250 (22.3)	273 (24.3)	3 (0.3)	1000 (89.2)	907 (80.8)
COVID-19	207 (25.3)	199 (24.3)	0	730 (89.2)	635 (77.6)
non-COVID-19	43 (14.1)	74 (24.3)	3 (1)	270 (88.8)	272 (89.5)
Fisher's exact test (p-values)					
<i>2019 vs 2021</i>	0.15	0.35	0.06	0.45	1.0×10^{-6}
<i>2019 vs 2021 non-COVID-19</i>	0.33	0.48	0.70	0.46	0.12
<i>2021 non-COVID-19 vs. COVID-19</i>	5.1×10^{-5}	1	0.02	0.83	3.5×10^{-5}

COI: conflicts of interest



High level of code sharing seen since [policy](#) was introduced requiring authors *“to make all author-generated code directly related to their study’s findings publicly available without access restriction at the time of publication unless specific legal or ethical restrictions prohibit public sharing of code.”*

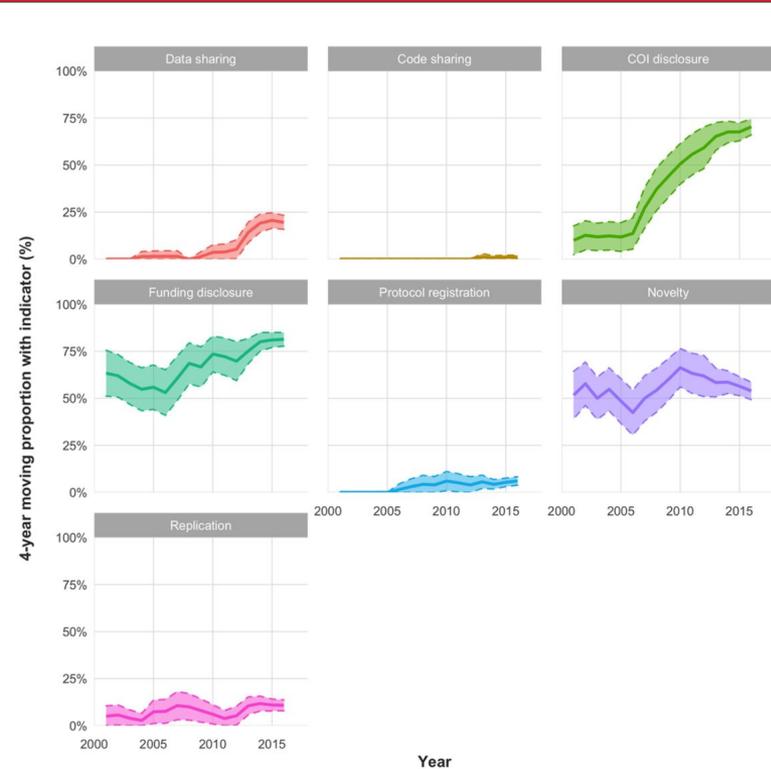
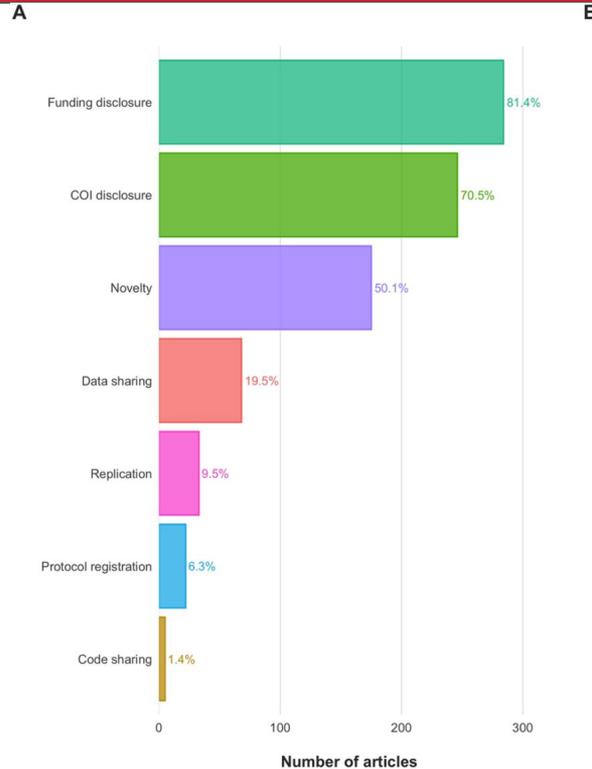


Code and data sharing in biomedical research



www.software.ac.uk

Code sharing in wider biomedical research still has a long way to go, although increasing, it is still low compared to other policies at 1.4% of articles in PubMed.



Data / software sharing among social researchers



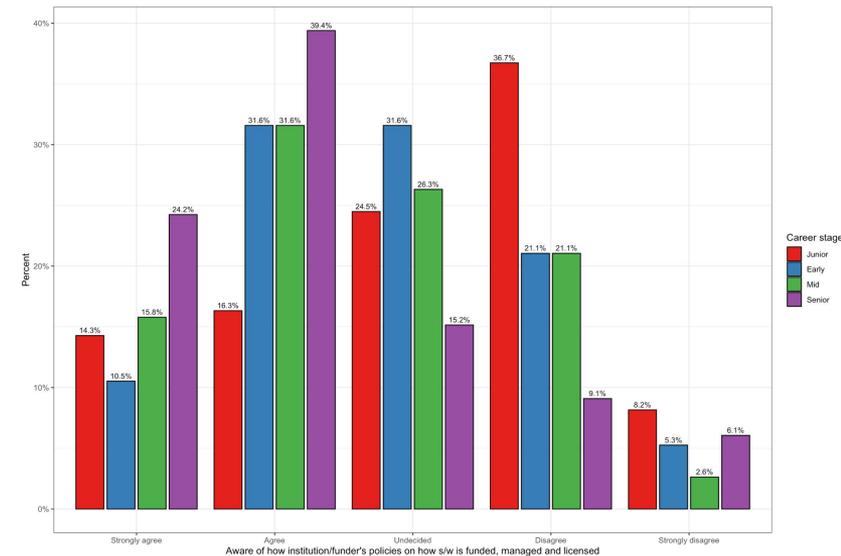
www.software.ac.uk

The UK Economic and Social Research Council policy is that *“all data created or repurposed during the lifetime of an ESRC grant must be made available for re-use or archiving within three months of the end of the grant.”*

Only 34% of 164 ESRC-funded researchers polled had shared data in the last five years.

Awareness of institutional / funder software policies varied by career stage

This matters, because 64% who shared data shared software widely, compared with 21% of those who didn't share data.



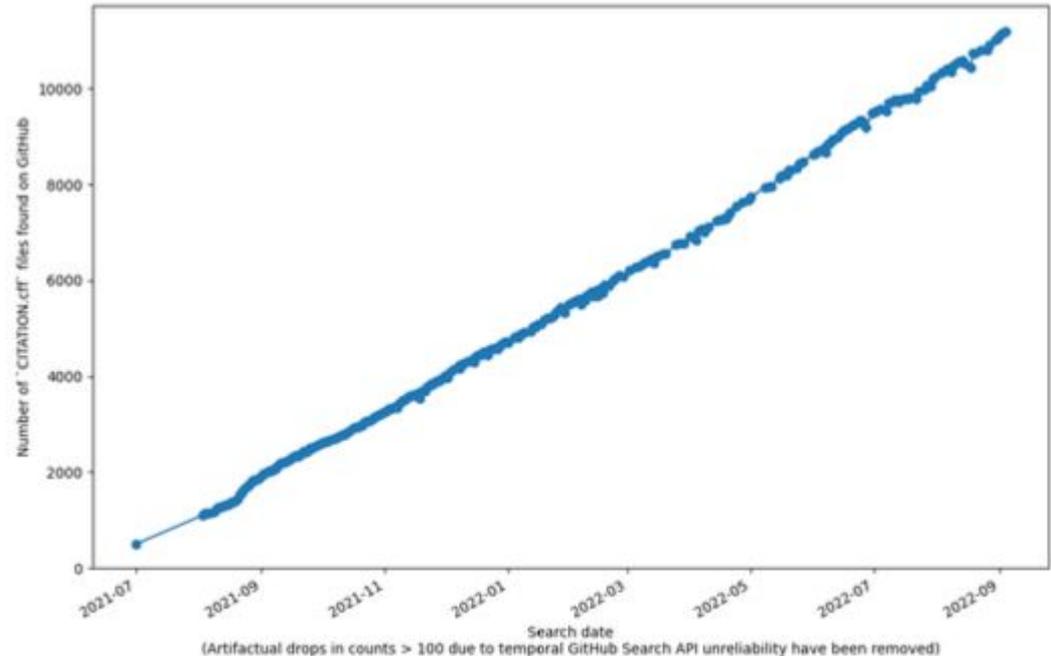
Software citation support in Github



www.software.ac.uk

GitHub code repository added support for [software citation using CFF files](#) in [July 2021](#).

By September 2022, over 11,000 project had added a citation files and the generated recommended citation information had been [viewed over 1.2m times](#).



Druskat, S., & Spaaks, J. H. (2022). The Citation File Format: Growing up to enable better software citation. Zenodo. <https://doi.org/10.5281/ZENODO.7049934>



The implementation and adoption of policy needs to overcome two types of barriers:

- Is it easy to translate the policy to a set of processes that work within the organisation / community?
- Will individuals in the organisation / community prioritise the application of these processes?
 - Because they're easy / cheap
 - Because there's a penalty for not doing
 - Because there's a benefit for doing

The more policies, the greater the risk of confusion or contradiction.

What are the key points in the research lifecycle where policies are effective?

Open research needs research software policy



www.software.ac.uk

Open research requires appropriate research software policy.

In general, adherence to research software and related policies is improving, as better guidance, tooling and support is provided by organisations.

However, there is still a long way to go, and care must be taken not to add excessive burden or confusion on researchers.

There are some key stakeholders who can help in implementation and adoption of policy including funders, publishers and research performing organisations.

Ultimately, it is down to researchers and research software engineers to follow the policies.