This is a draft output for the joint Research Software Alliance and FORCE11 Code Availability taskforce. To sign up to work or comment on the draft, please contact code-availability+owner@googlegroups.com

Publisher Policies on Code: A High Level Landscape Analysis

Written by Lauren Cadwallader, Tom Honeyman, Hugh Shanahan

Background

At the 19th Research Data Alliance Plenary (June 2022), a Birds of a Feather session was convened to discuss aspects of code sharing and investigate if there was a consensus on which aspect of code sharing the community should be a focus of attention. The group, which included publishers, institutions, researchers and infrastructure providers, agreed to focus on “code” (i.e., “single use” or ”one-shot”, often referred to as “analysis code”) that usually accompanies published articles and increasing the availability of this code. Note that this refers to a specific context in which “source code” is created and not “code” more generally.
Strategies to increase adoption of code sharing include policies, tools, infrastructure and guidance. As an initial starting point, we have carried out a high level landscape analysis of publisher policies around code sharing. This will allow us to evaluate the current expectations for authors in general across the scholarly landscape and target future action at the appropriate level. We hope to present a more in depth landscape analysis at the 20th Research Data Alliance Plenary (March 2023).

Methods

To restrict the total number and variability of policies being analysed, the initial focus was on publisher-wide policies as opposed to journal specific author guidelines.

The CHORUS Publisher Data Availability Policies Index was used to define a list of publishers for this exercise. Forty-eight publishers were identified via this list, although some publishers belong to the same company, e.g. F1000 and Taylor & Francis. Each publisher’s webpages were checked for details of a policy on code sharing and the degree to which they required code to be shared (e.g. required versus encouraged). Excerpts of relevant policy were extracted where possible, or summarised when the detail was dispersed over a large amount of text.

For the purposes of this high level summary, the policies were coded based on a limited number of features. A more in depth assessment will be conducted for RDA in March 2023.
The policies were coded according to the following distinctions with the possible answers given in square brackets:

- Type of policy on code sharing [**no policy**/**encouraged/required**]

- Should code be mentioned in a Data Availability Statement (or equivalent)? This is regardless of the type of policy [**Yes**/**No**]
- Is the code required to be in a repository? [**Yes**/**No**]
- Is software citation mentioned in the policy or related text? [**Yes**/**No**]

Results

Of the 36 publishers assessed from the CHORUS list, 56% (n=20) had no policy on code sharing, 33% (n=12) encouraged sharing of code and 11% (n=4) required code sharing. 8% of the sample (n=3) required code to be deposited into a repository.
Just over a third (36%) (n=13) encouraged or required the location of any shared code to be detailed in an Availability Statement and only 28% (n=10) encouraged or required code citation in their policy.

One policy (3%) encouraged or required the code location in Availability Statements and code citation as well as required repository use. If repository use is downgraded from required to encouraged, this number rises to 5. Policies encouraging code sharing generally discussed these features more often than policies requiring code sharing. Where they are mentioned repositories are more often encouraged than required and the same is true of availability statements (see Table 1).

Table 1. Intersection of features - repository and availability statements - and the strength of the policy in relation to the feature.

	Repository not mentioned	Repository encouraged	Repository required
Availability Statement not mentioned	15	3	0
Availability Statement encouraged	2	5	0
Availability Statement required	0	1	2

Discussion

This initial landscape analysis has highlighted a number of key points about code sharing policies in scholarly publishing, which may be potential future avenues of work:

Prevalence of policies

Far fewer publishers have code sharing policies in comparison to data sharing policies. Of the CHORUS list sampled, 44% had a code sharing policy whereas 92% have a data sharing policy. This is potentially slightly skewed as some publishers have overarching data policies but journal specific code sharing policies but nevertheless shows a gulf between how these two research outputs are treated. A better understanding of the prevalence of code sharing policies will allow us to highlight this shortfall to publishers and be able to pitch suggested templates at an appropriate level acknowledging that different publishers are at different stages of code sharing.

Inconsistency in requirements within publishers

Not all publishers use a standard code sharing policy across all their journals. This may be for a variety of reasons, such as that a standard code sharing policy is not appropriate for all journals or the journal may be run by a Society who determines their own journal policies. This is also the case for data sharing policies and work has been done to try to standardise the range of data sharing policies across journals/publishers (Hrynaszkiewicz et al 2020).

Non-publisher-wide code sharing policies make it difficult to track the presence of policies but at least with the adoption of tiered data sharing policies, such as those implemented by Springer Nature and Taylor & Francis, there is a central point of information on data sharing at the publisher. This approach can then account for disciplinary or journal differences in readiness for sharing policies (Jones et al 2019). Highlighting this inconsistency will help make the case for the need for publisher alignment (both within and between publishing organisations), which simplify the support publishers offer authors and can bring operational and infrastructural benefits.

Inconsistency in requirements for code sharing

In addition to the inconsistency within publishers, there is also inconsistency between publishers with respect to code (and data) sharing policies. Different publishers require, or encourage, different aspects of code sharing which makes it complex for researchers, and research support staff, to understand and comply with as well as challenging for infrastructure providers (Hrynaszkiewicz et al. 2020). This high level analysis points to a varied landscape of code policies, although the full extent of the fragmentation of policies will only be known when further analysis is undertaken to review requirements around features such as persistent identifiers, archival copies of code, containerisation etc. Reducing the fragmentation of policies and seeking alignment between publishers will help reduce the burden faced by authors who have to navigate publisher policies. In addition, this will aid research support staff at institutions who have to offer support to researchers as it will also reduce the burden they face.

Inconsistent reference to code

Another aspect that is inconsistent is how or where code sharing is indicated in publisher editorial policies. Some publishers treat code as a standalone research output with its own policy. Other publishers include it in their definition of research data (see Royal Society of Chemistry 2022 as an example). Often in these cases the policy name doesn't reflect the breadth of outputs that could be in scope, which risks the discoverability of these requirements as it requires authors to take in, understand and process the entirety of the policy. An example of better practice is the Wellcome Trust's "Data, software and materials management and

sharing policy", which is clear from the title that several output types are covered in the policy. Ensuring code is treated as a research output separate from data or software when it comes to policies would increase its visibility and make it easier for authors, reviewers and editorial to find and understand what is required from the author.

Next steps

We would like to propose that this group conducts a fuller analysis of publisher policies into code sharing. This will enable us to drill down into the features that are present (or not) and help us understand the publisher landscape in relation to code sharing policies. The fuller analysis will assess:

- If sharing of code is {not suggested, encouraged, mandated}
- Where it will be available is {not suggested, suggested, mandated}
- Where is will be available includes {record preservation, trusted repository, version control, persistent identifiers, non persistent storage / user editable, enforced metadata standards} (inclusive, not all suggested infrastructure needs to demonstrate it)
- If sharing availability statements are {not suggested, suggested, mandated}

Publishers on the CHORUS will be used for this assessment as well as those belonging to the STM Association. To the extent that it is practical we will extend to an analysis of journal policies as surfaced through the work of the TOP Factor analysis.

On the basis of the fuller coding, we will identify the best targets to approach to enter into discussion about policy, with suggestions of more consistent wording, and an analysis of their relative standing amongst other publishers, and through an assessment of journal policies within publishers

Data Availability

The data gathered for this exercise is currently available as a GSheet. Data from the proposed fuller analysis will be shared in a repository.

Bibliography

Hrynaszkiewicz, Iain; Simons, Natasha; Hussain, Azhar; Grant, Rebecca; Goudie, Simon (2020): Developing a research data policy framework for all journals and publishers. CODATA Data Science Journal. http://doi.org/10.5334/dsj-2020-017

Jones L, Grant R, Hrynaszkiewicz I. 2019. Implementing publisher policies that inform, support and encourage authors to share data: two case studies. Insights 32:11. DOI: 10.1629/uksg.463.

Royal Society of Chemistry (2022) Data sharing. Guidance for best practice and reproducibility of experimental data [webpage, accessed 17 Oct 2022]

https://www.rsc.org/journals-books-databases/author-and-reviewer-hub/authors-information/prepare-and-format/data-sharing/