

Getting credit for your data: data citation and publication

Sarah Callaghan (NCAS-BADC) and the NERC Data Citation and Publication team
sarah.callaghan@stfc.ac.uk

Why should I publish and cite data?

It's good for science:

- Allows research to be reproduced and verified.
- Allows data to be found and understood by other researchers.
- Encourages collaboration and cross-disciplinary work.
- Maintains the scientific record.

It's good for researchers:

- Provides attribution and credit for the hard work of creating a dataset.
- Makes it easier to build in good data management practices.
- Demonstrates the quality and usefulness of the data (especially for researchers outside the immediate field).
- Captures information about the data's impact.

It's good for funders:

- Reduces money and effort needed to recreate important datasets.
- Reassures funders that the data resulting from their funding is useful and archived safely.
- Captures information about the funder's impact.

It's good for data repositories:

- Encourages researchers to deposit data where it can be archived and curated.
- Papers linked to the data are an excellent source of metadata.
- Captures information about the repository's impact.

You've convinced me. How do I cite data?

- Treat it like a journal article.
- If you've used data in your research, and mention it in your paper, cite it in the text (Callaghan *et al*, 2009) and put the full citation in the paper's reference list.
- Journals should have guidance on data citation for authors. If they don't, ask!

But how do I know what the right data citation is?

- The data repository should tell you how to cite the dataset on the dataset catalogue page. If they don't – ask them!

- If in doubt, use the DataCite recommended citation format:
 - Creator(PublicationYear): Title. Publisher. Identifier

Acknowledgements: The NERC data citation and publication project is funded through the NERC Science Information Strategy programme.

A journal editor has told me I need a DOI for the data discussed in my paper. How do I get one?

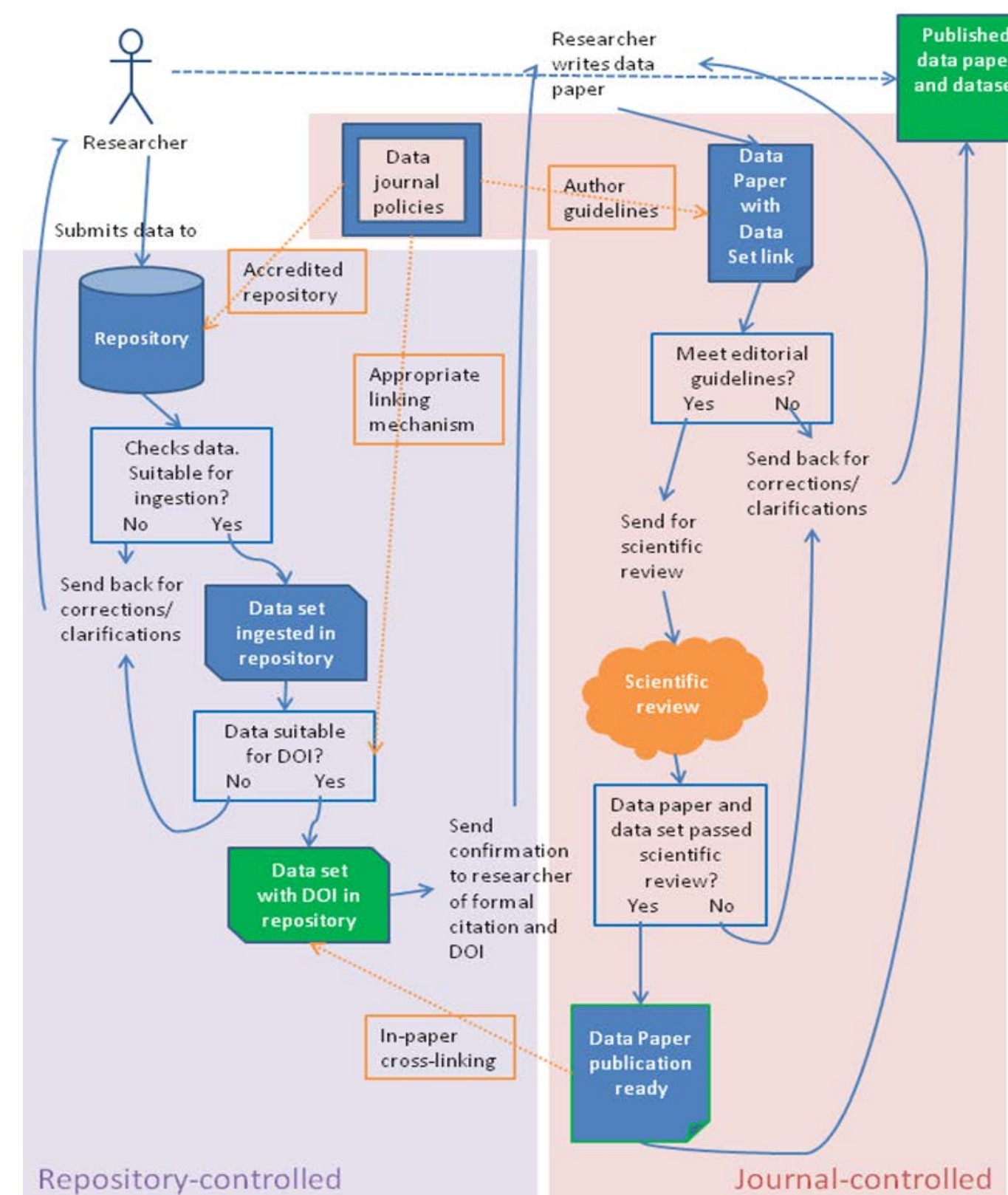
- If your data **is** already ingested into BADC, then put in an email request for a DOI by emailing support@ceda.ac.uk. We may come back to you to ask questions about the dataset, mainly about how finished it is, and to check the important bits of the metadata that become the citation. If all is well, we'll then mint the DOI for you.
- If your data **isn't** already ingested into BADC (or other NERC repository), email us at support@ceda.ac.uk anyway. We'll either work with you to get the data ingested and DOI minted, or will suggest other options. Ingesting the data may take some time, so be prepared!

I have a dataset I'm proud of. How do I publish it?

1. Find a suitable data journal (for example, *Geoscience Data Journal*, published by Wiley, *Earth Systems Science Data*, published by Copernicus, or *Scientific Data*, published by Nature)
2. Follow their guidelines and processes for data publication

Can you generalise the process for me?

Have a look at this flowchart:



The researcher writes a data article which describes the dataset, giving details of its collection, processing, software, file formats, etc., without needing to do novel analyses or come up with ground breaking conclusions. The data article explicitly cites the dataset the article is about. Preferably the dataset itself is already ingested into a trusted repository that can provide permanent identifiers (such as DOIs). If not, that can be done while writing the data article.

The data article then gets submitted using the data journal's online submission process. A lot of stuff happens behind the scenes, done by the repository and the journal, including peer-review of the article and the data.

Then the article gets published on-line, with a link back to the dataset.

Dataset information and link

I still have questions!

Email the author at sarah.callaghan@stfc.ac.uk, or support@ceda.ac.uk
We're always happy to help!