

DOI: Digital attachments for Scientific Papers



Ever saw a claim on a paper you disagreed with or got triggered by, and then wanted to reproduce the experiment? Good luck finding the code and the data used in the experiments.

When we want to redo experiments of papers, it starts with finding the code and data used. A good start is Github or the homepage of the scientist. Also Gitlab, Bitbucket, SourceForge or the personal homepage of one of the researchers could be a place to look. Emailing the authors is often only an option, if the university homepage mentions such option - we're not surprised to get no reaction at all. If all that doesn't work, then implementing the pseudo-code and creating own data might be the only option - not if that will support the claims.

So what if scientific papers had an easy way to connect to digital objects like code and data?

Here the DOI comes in.

DOI: Digital Object Identifier

A Digital Object Identifier (DOI) is a persistent identifier or handle used to uniquely identify objects, standardized by the International Organization for Standardization (ISO). Organisations which work with DOIs, provide a library where each entry is identified by unique ID. As the DOI is a url starting with <http://dx.doi.org>, it brings you directly to the sources.

The European organisation ILL is very happy with the results and is promoting its usage. See here the video they made:

YouTube Video: [YouTube.com/watch?v=40B28q8EUyI](https://www.youtube.com/watch?v=40B28q8EUyI)

[\(link\)](#)

Not only Europe is working more with DOIs. It's a truly international organisation with users worldwide, from East to West.

Wait, DOI is not only for linking to the PDF?

Most usage of DOI is for a link to exactly the same PDF. And here is the current problem: the solution exists, but is not fully used.

These pages should have links to code and data too. The good part is that many websites like IEEE Xplore and ACM digital library already have the DOI on their papers, just missing the code and data.

How to share research data yourself?

Step 1: find a good license for code and data

Avoiding a situation where commercial companies run away with your code, a dual license is appropriate. It's out of scope for this

blog, so check [Multi-license on Wikipedia](#) to start searching what is best for your code. If you do want to commercialise your code, think of partnering with us and get in [contact](#) before publishing.

Step 2: get a DOI number

To add a DOI, there is much information found on the [DOI website](#), including a [handbook](#) and a [list of registrars](#). As it needs to be centrally administrated per institute or publisher, it's best to look what your colleagues have done.

Step 3: publish data and code

To add data and code, it's important the links don't disappear after a career switch. This is the hard part, as not all universities and institutes have a public repository for code and download-center for data. See for example [what NASA Earthdata has put in place](#), but this will be different for many other organisations.

Step 4: Get more citations

Seeing is believing. In a time where the label "fake news" is often used, it is much more important to seriously support what you claim. It also helps in having others doing more benchmarks on different hardware, while citing your research.

While at it, why not update your past papers?