

# Scientific results traceability: software citation using GitHub and Zenodo

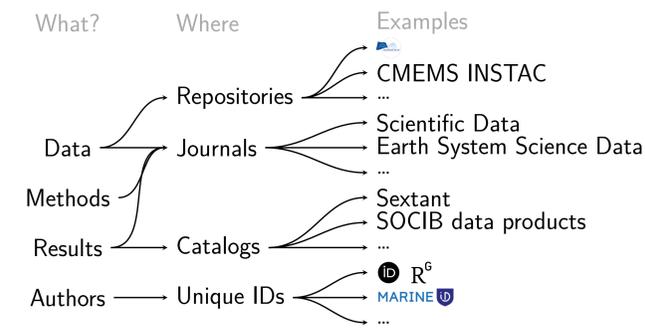
Oceanography
 Data analysis
 Data citation
 Reproducibility
 SeaDataCloud
 ODIP
 Python
 Jupyter
 GitHub
 Zenodo
 orcid

## What will we talk about ?

- DOI** Digital Object Identifier: unique alphanumeric string assigned to identify content and provide a persistent link to its location on the Internet.
- GitHub**: web-based hosting service for version control using git <http://github.com/>
- ORCID**: persistent digital identifier to distinguishes researchers <https://orcid.org/>
- Zenodo**: a repository to deposit scientific papers and/or research data <https://zenodo.org/>

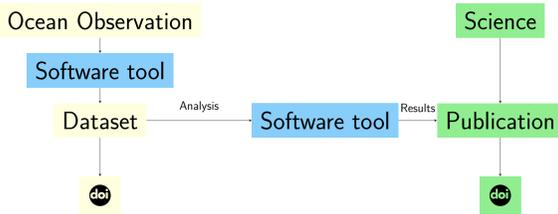
## How can we foster reproducibility ?

By making  $\left\{ \begin{array}{l} \text{Data} \\ \text{Method} \\ \text{Results} \end{array} \right\}$  available and citable

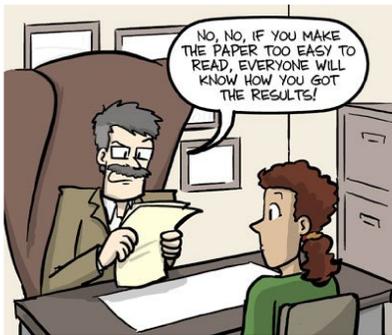


## Unique identifiers for each components

- DOI** – minted on  $\left\{ \begin{array}{l} \text{Datasets} \\ \text{Data products (e.g., climatologies)} \end{array} \right\}$
- ID** – identifies authors



## Is this sufficient to go from data to results ?



Source: "English Communication for Scientists", Nature.

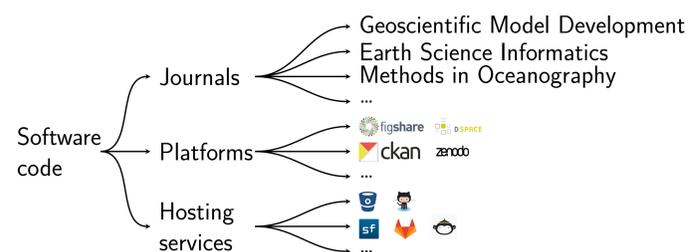
<https://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993/writing-scientific-papers-14239285>

The software tools should also be properly versioned and cited to document:

- how the observations are converted to a dataset,
- how scientific results are derived from the dataset.

## Making the code available

If the code of the software tool(s) is open, there are several ways to distribute it:



## Research platforms

Online infrastructures whose objective is to persistently store and archive digital artifacts relevant to research: articles, data, images, code, model outputs, ...

Comparison

| Tool         | CKAN                | DSpace        | Figshare     | Zenodo   |
|--------------|---------------------|---------------|--------------|----------|
| Open Source  | Yes                 | Yes           | No           | Yes      |
| Licence      | Affero GNU GPL v3.0 | BSD           | -            | GPL-2.0  |
| 1st released | November 2011       | November 2002 | January 2011 | May 2013 |
| Language     | Python              | Java          | -            | Python   |
| Deployment   | Local               | Local         | Cloud        | Cloud    |
| with GitHub  | No                  | No            | Yes          | Yes      |
| with ORCID   | Yes                 | Not direct    | Yes          | Login    |

## Let's have a closer look at Zenodo

- ✓ ingest all research outputs and any file format
- ✓ DOIs assigned to have uniquely citable files
- ✓ integrated into reporting lines for research via OpenAIRE.

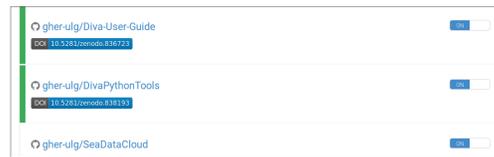
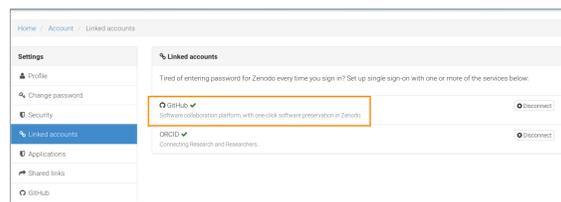
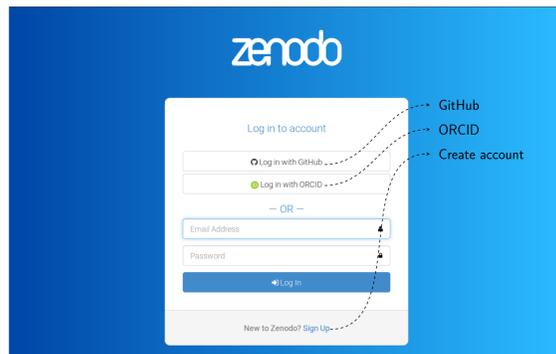


Figure 1: Zenodo log in: using other accounts is great feature. When linked with GitHub, Zenodo creates a DOI every time one makes a release in one of the enabled repositories.

We now have all the pieces to cite the code used in the research:

- Zenodo login using GitHub or ORCID
- Upload of software code to GitHub or to Zenodo
- Generation of the DOI for a given version of the code

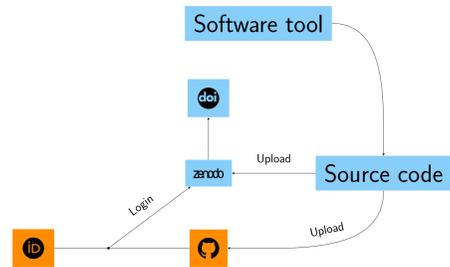


Figure 2: Getting unique identifiers for software codes using Zenodo.

## Example 1

From observations to dataset.

The SOCIB glider toolbox is a set of MATLAB/Octave scripts to manage data collected by a glider fleet: data download, processing and figure generations.

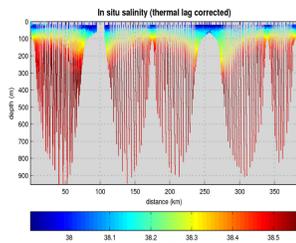


Figure 3: Example of a glider section produced by the toolbox.

The code development was carried out in GitHub ([https://github.com/socib/glider\\_toolbox](https://github.com/socib/glider_toolbox), credits to T. Garau and J.P. Beltran) and was recently coupled to Zenodo.

DOI [10.5281/zenodo.836706](https://doi.org/10.5281/zenodo.836706)

## Example 2: the DIVA interpolation tool

1990's: Variational Interpolation Method (Fortran 77) only 2D interpolations

2006 SeaDataNet, code refactory and set of bash scripts

2007 with ODV

2008 code in Subversion, distribution through GHER web page

2009 new modules in Fortran 90 for loops over depth and time

2012 new error calculation technique

2017 new version control system:

- Switch from SVN to git, distribution via GitHub, sync with zenodo
- Enable Diva repository in Zenodo
- Edit the different tags on GitHub to get DOI

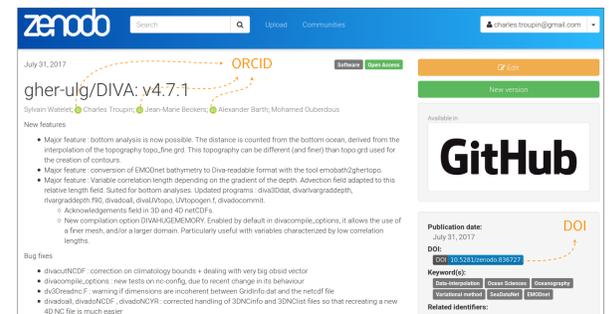


Figure 4: Main page of DIVA software in Zenodo. Note the ORCID logo with the authors and the DOI relative to the code. Share on social media and "cite as" options.

## Putting all the pieces together

To ensure reproducibility and traceability, unique identifiers (DOI) are attributed to:

- Datasets
- Software tools
- Authors
- Scientific results

Ideally, all the identifiers should be present in the published version of the research paper.

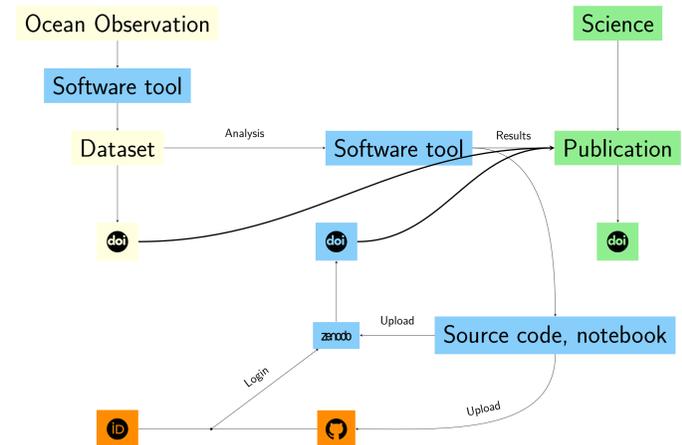


Figure 5: From data to final results: all the components are identified and citable.

## Acknowledgements

The work presented in this poster has been developed in the frame of SeaDataCloud—Further developing the pan-European infrastructure for marine and ocean data management, Project ID 730960 and ODIP 2—Extending the Ocean Data Interoperability Platform, Project ID: 654310.

