C. Troupin*, C. Muñoz*,
M. A. Rújula*, J. G. Fernández*,
S. Watelet*, A. Barth* & J.-M. Beckers*

*GHER-University of Liège
*Balearic Islands Coastal Ocean
Observing and Forecasting System

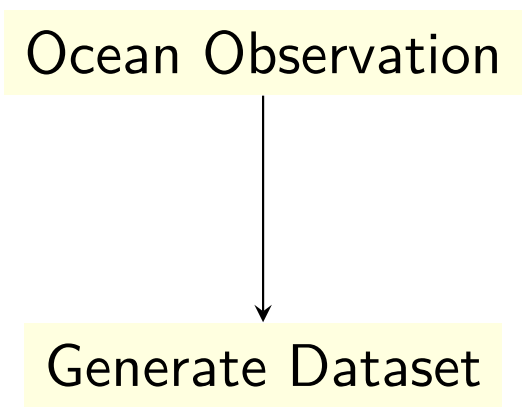
# *Software citation & process traceability*

# Persistent identifiers everywhere

Ocean Observation

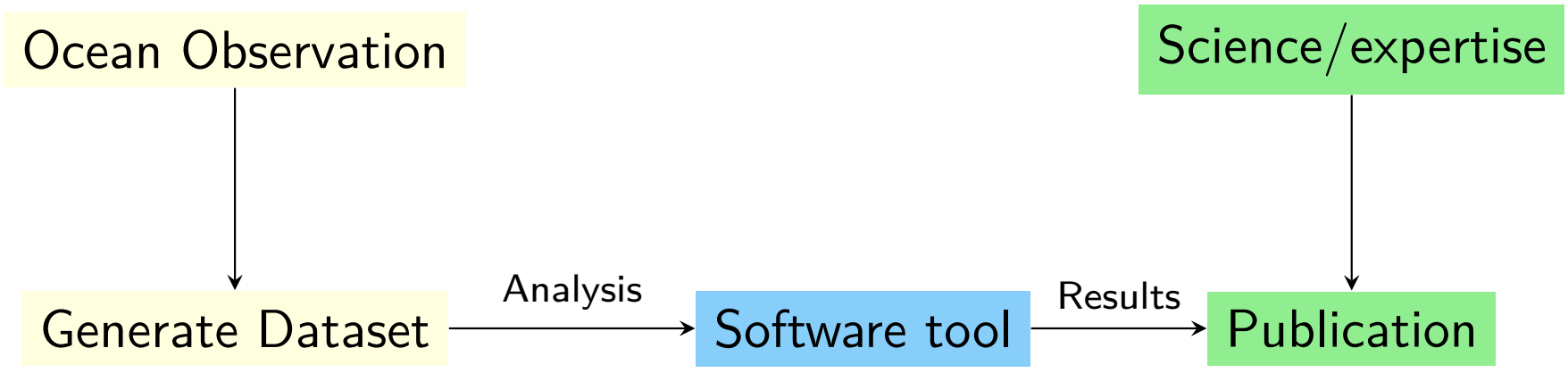
Science/expertise

# Persistent identifiers everywhere

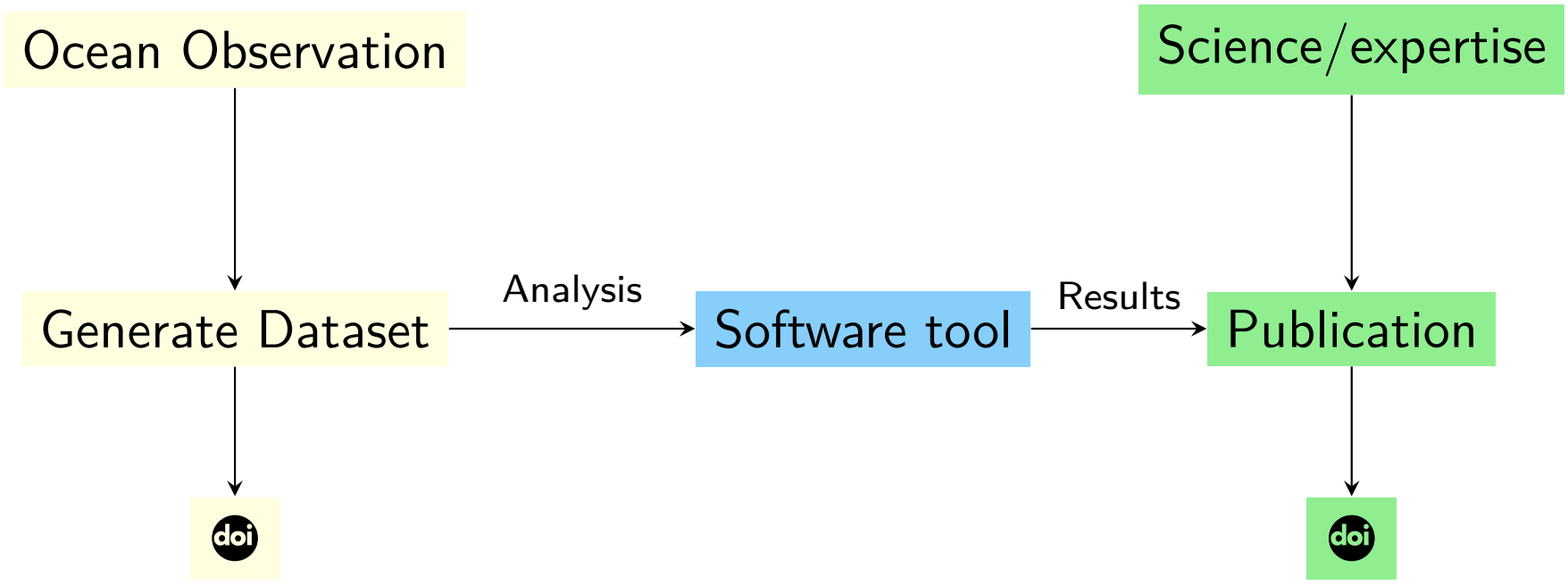



Science/expertise

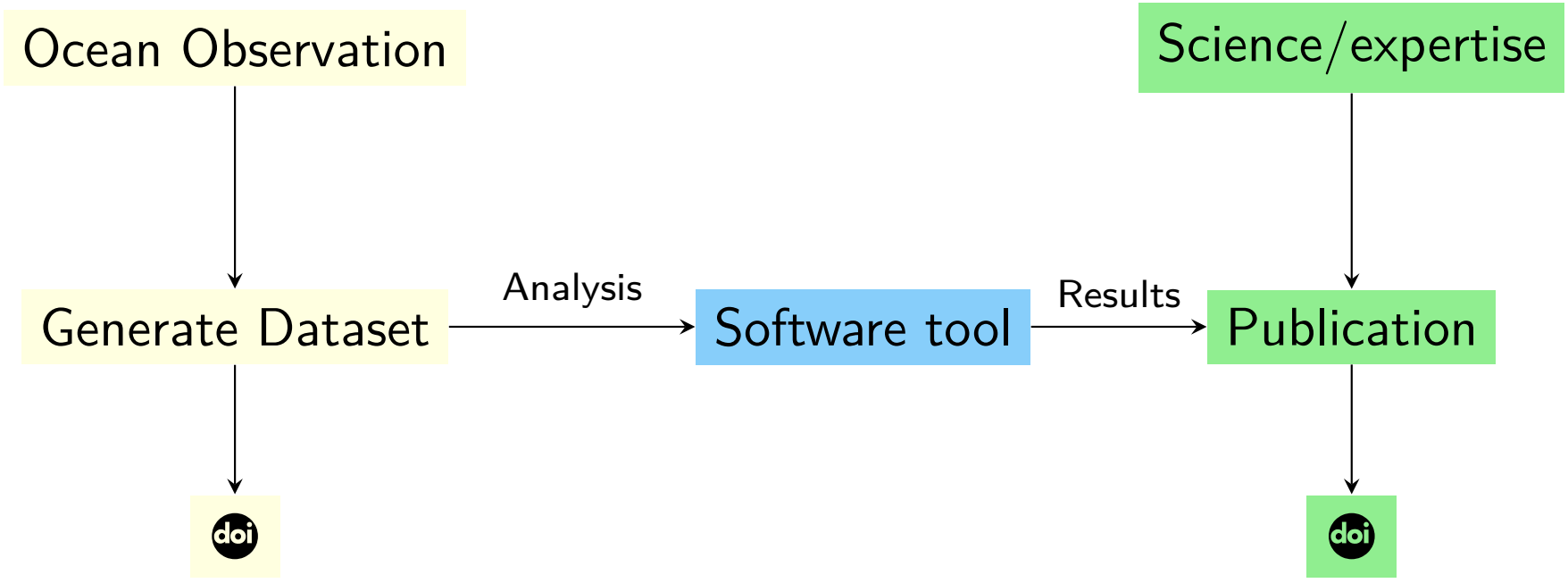
# Persistent identifiers everywhere



# Persistent identifiers everywhere



# Persistent identifiers everywhere

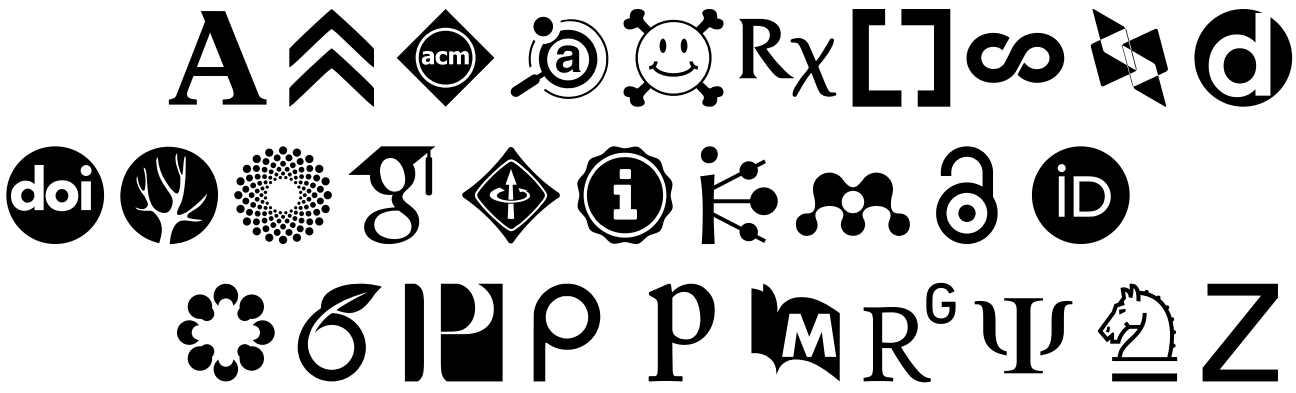



How can we ensure the readers/users can reproduce the results?

Persistent identifiers:
what about

*software tools?*

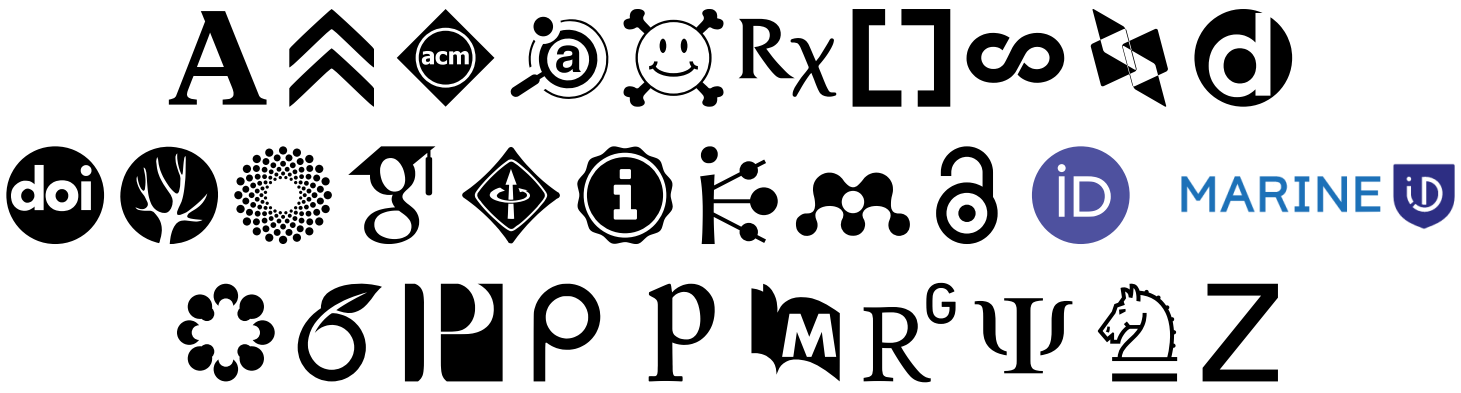
# Context: who has done what, and how?




Source: Academicons

# Context: who has done what, and how?

Let's work with ORCID and MarineID
(see previous ODIP workshops)

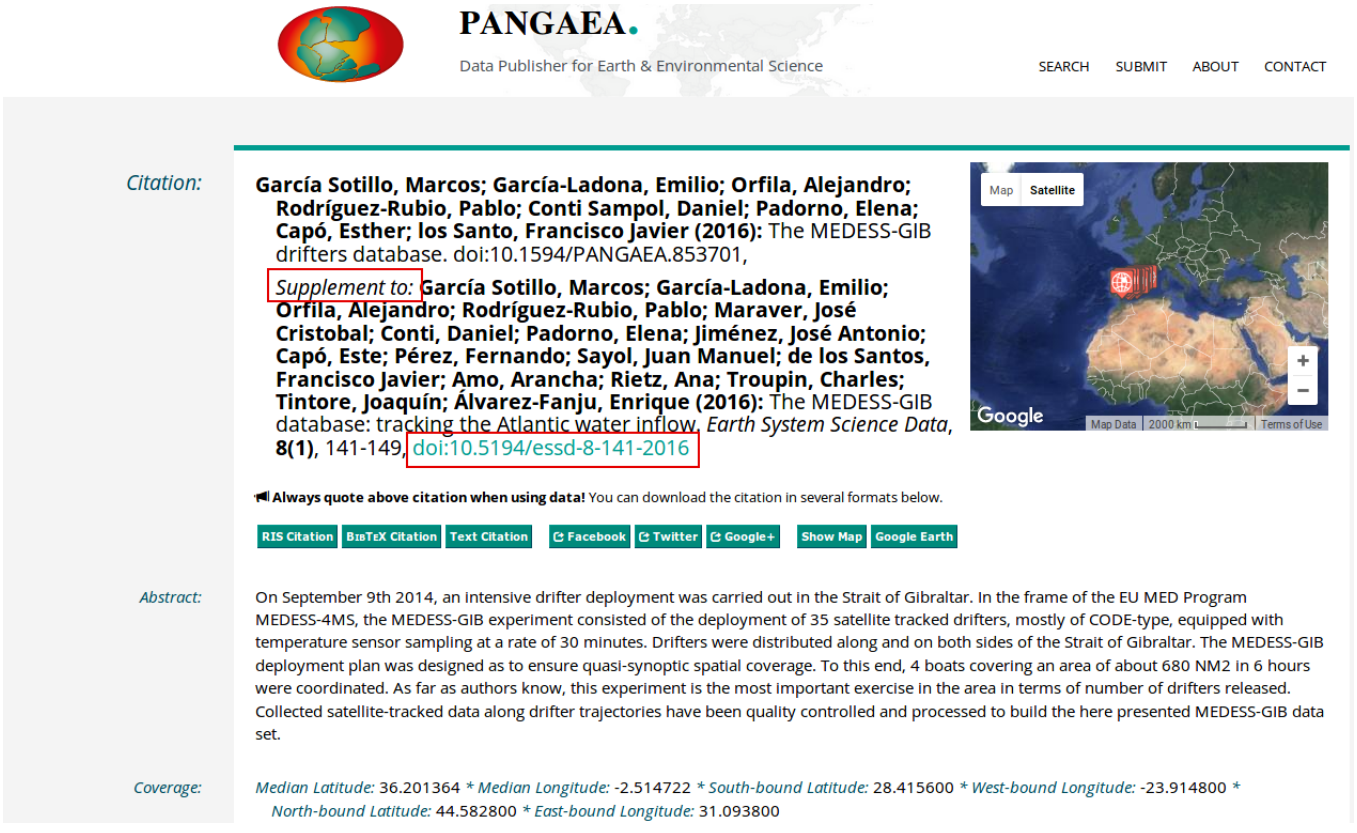



Source: Academicons

# Data set identification and citation

See previous ODIP workshops + links to other initiatives

- Research Data Alliance Data Citation working group
- THOR project
- Pangaea



PANGAEA.

Data Publisher for Earth & Environmental Science

SEARCH SUBMIT ABOUT CONTACT

*Citation:*

**García Sotillo, Marcos; García-Ladona, Emilio; Orfila, Alejandro; Rodríguez-Rubio, Pablo; Conti Sampol, Daniel; Padorno, Elena; Capó, Esther; los Santo, Francisco Javier (2016):** The MEDESS-GIB drifters database. doi:10.1594/PANGAEA.853701,

*Supplement to:* **García Sotillo, Marcos; García-Ladona, Emilio; Orfila, Alejandro; Rodríguez-Rubio, Pablo; Maraver, José Cristobal; Conti, Daniel; Padorno, Elena; Jiménez, José Antonio; Capó, Este; Pérez, Fernando; Sayol, Juan Manuel; de los Santos, Francisco Javier; Amo, Arancha; Rietz, Ana; Troupin, Charles; Tintore, Joaquín; Álvarez-Fanju, Enrique (2016):** The MEDESS-GIB database: tracking the Atlantic water inflow, *Earth System Science Data*, **8(1)**, 141-149, doi:10.5194/essd-8-141-2016

Map Satellite

Google Map Data 2000 km Terms of Use

**Always quote above citation when using data!** You can download the citation in several formats below.

RIS Citation | BibTeX Citation | Text Citation | Facebook | Twitter | Google+ | Show Map | Google Earth

*Abstract:*

On September 9th 2014, an intensive drifter deployment was carried out in the Strait of Gibraltar. In the frame of the EU MED Program MEDESS-4MS, the MEDESS-GIB experiment consisted of the deployment of 35 satellite tracked drifters, mostly of CODE-type, equipped with temperature sensor sampling at a rate of 30 minutes. Drifters were distributed along and on both sides of the Strait of Gibraltar. The MEDESS-GIB deployment plan was designed as to ensure quasi-synoptic spatial coverage. To this end, 4 boats covering an area of about 680 NM2 in 6 hours were coordinated. As far as authors know, this experiment is the most important exercise in the area in terms of number of drifters released. Collected satellite-tracked data along drifter trajectories have been quality controlled and processed to build the here presented MEDESS-GIB data set.

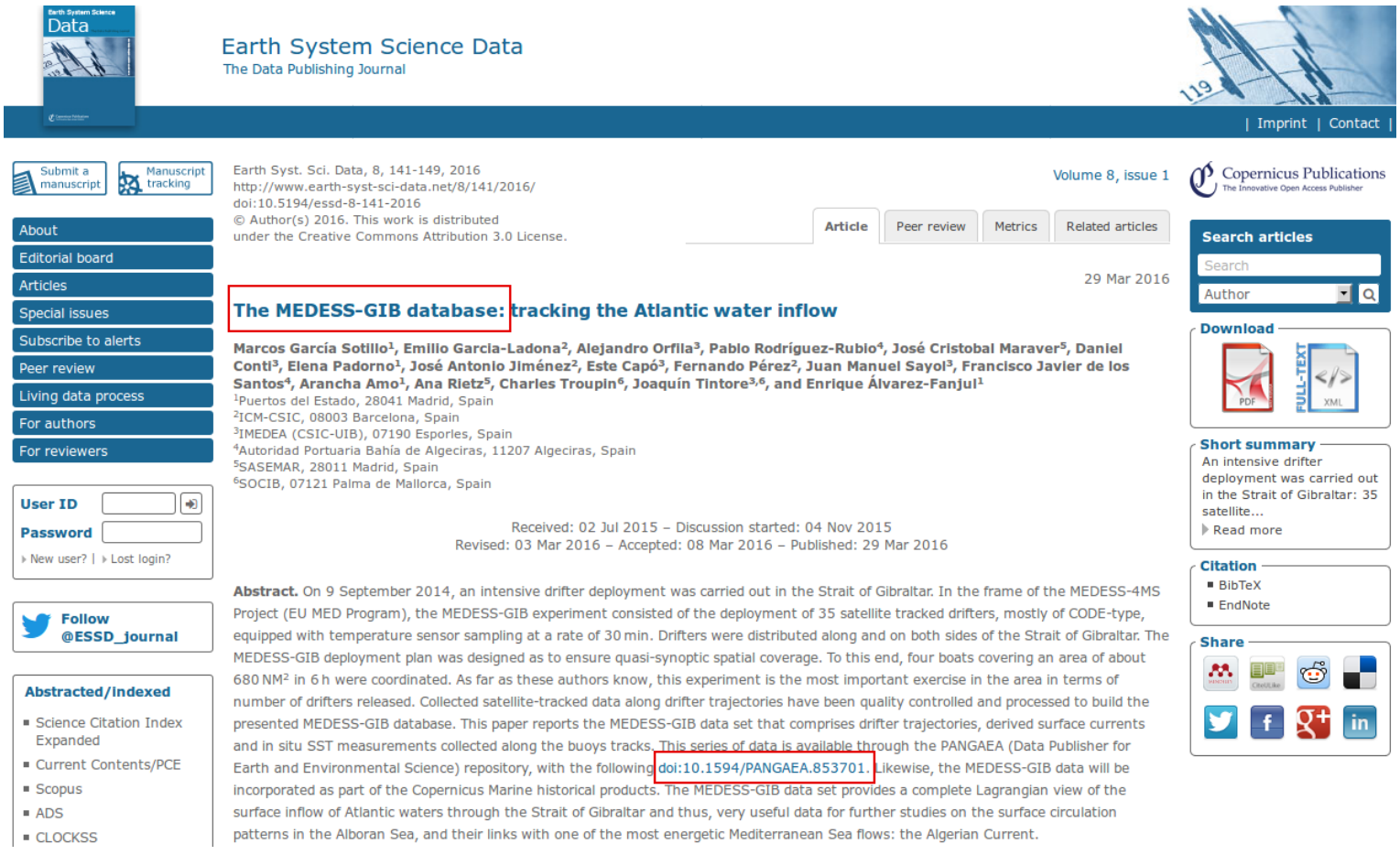
*Coverage:*

*Median Latitude:* 36.201364 * *Median Longitude:* -2.514722 * *South-bound Latitude:* 28.415600 * *West-bound Longitude:* -23.914800 * *North-bound Latitude:* 44.582800 * *East-bound Longitude:* 31.093800

# Data sets in peer-reviewed journals

## Earth System Science Data

- "*reuse of high-quality data of benefit to Earth system sciences*"
- 30 articles in 2017 (as of July 24th)



Earth System Science Data
The Data Publishing Journal

| Imprint | Contact |

Submit a manuscript | Manuscript tracking

About
Editorial board
Articles
Special issues
Subscribe to alerts
Peer review
Living data process
For authors
For reviewers

User ID
Password
New user? | Lost login?

Follow @ESSD_journal

Abstracted/Indexed
- Science Citation Index Expanded
- Current Contents/PCE
- Scopus
- ADS
- CLOCKSS

Earth Syst. Sci. Data, 8, 141-149, 2016
http://www.earth-syst-sci-data.net/8/141/2016/
doi:10.5194/essd-8-141-2016


Volume 8, issue 1

Article | Peer review | Metrics | Related articles

29 Mar 2016

# The MEDESS-GIB database: tracking the Atlantic water inflow

**Marcos García Sotillo[1], Emilio Garcia-Ladona[2], Alejandro Orfila[3], Pablo Rodríguez-Rubio[4], José Cristobal Maraver[5], Daniel Conti[3], Elena Padorno[1], José Antonio Jiménez[2], Este Capó[3], Fernando Pérez[2], Juan Manuel Sayol[3], Francisco Javier de los Santos[4], Arancha Amo[1], Ana Rietz[5], Charles Troupin[6], Joaquín Tintoré[3,6], and Enrique Álvarez-Fanjul[1]**

[1]Puertos del Estado, 28041 Madrid, Spain
[2]ICM-CSIC, 08003 Barcelona, Spain
[3]IMEDEA (CSIC-UIB), 07190 Esporles, Spain
[4]Autoridad Portuaria Bahía de Algeciras, 11207 Algeciras, Spain
[5]SASEMAR, 28011 Madrid, Spain
[6]SOCIB, 07121 Palma de Mallorca, Spain

Received: 02 Jul 2015 – Discussion started: 04 Nov 2015
Revised: 03 Mar 2016 – Accepted: 08 Mar 2016 – Published: 29 Mar 2016

**Abstract.** On 9 September 2014, an intensive drifter deployment was carried out in the Strait of Gibraltar. In the frame of the MEDESS-4MS Project (EU MED Program), the MEDESS-GIB experiment consisted of the deployment of 35 satellite tracked drifters, mostly of CODE-type, equipped with temperature sensor sampling at a rate of 30 min. Drifters were distributed along and on both sides of the Strait of Gibraltar. The MEDESS-GIB deployment plan was designed as to ensure quasi-synoptic spatial coverage. To this end, four boats covering an area of about 680 NM$^2$ in 6 h were coordinated. As far as these authors know, this experiment is the most important exercise in the area in terms of number of drifters released. Collected satellite-tracked data along drifter trajectories have been quality controlled and processed to build the presented MEDESS-GIB database. This paper reports the MEDESS-GIB data set that comprises drifter trajectories, derived surface currents and in situ SST measurements collected along the buoys tracks. This series of data is available through the PANGAEA (Data Publisher for Earth and Environmental Science) repository, with the following doi:10.1594/PANGAEA.853701. Likewise, the MEDESS-GIB data will be incorporated as part of the Copernicus Marine historical products. The MEDESS-GIB data set provides a complete Lagrangian view of the surface inflow of Atlantic waters through the Strait of Gibraltar and thus, very useful data for further studies on the surface circulation patterns in the Alboran Sea, and their links with one of the most energetic Mediterranean Sea flows: the Algerian Current.

Copernicus Publications
The Innovative Open Access Publisher

Search articles
Search
Author

Download
PDF | FULL-TEXT XML

Short summary
An intensive drifter deployment was carried out in the Strait of Gibraltar: 35 satellite...
Read more

Citation
- BibTeX
- EndNote

Share

# Data sets in peer-reviewed journals

## Scientific Data

- "*promote wider data sharing and reuse, and to credit those that share*"
- 127 publications in 2017 as of September 4 (all disciplines)

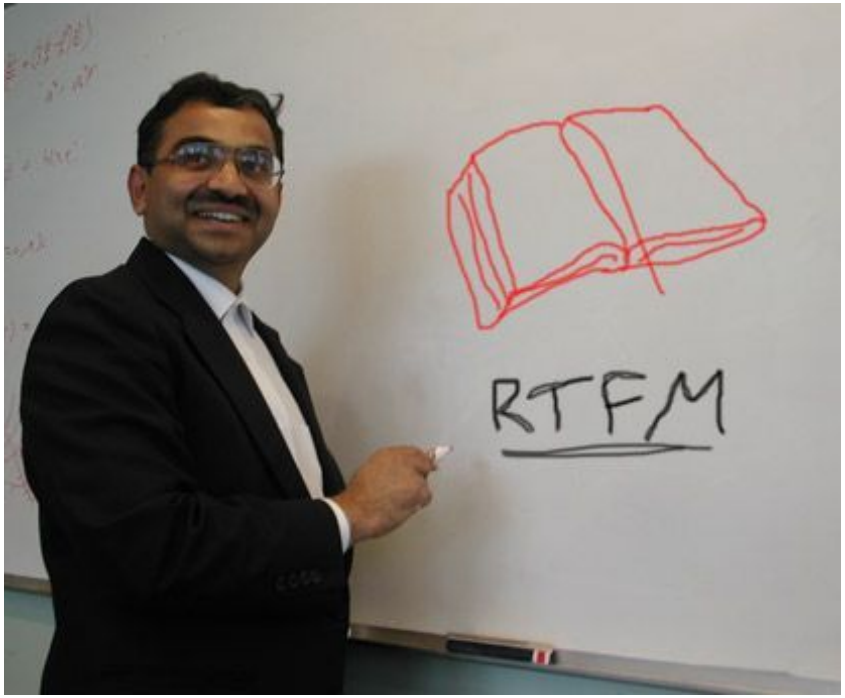
# How to go from data to products?

# How to go from data to products?

- Read the publication?

# How to go from data to products?

- Read the publication?
- Read the manual?




# How to go from data to products?

- Read the publication?
- Read the manual?
- Get and re-use code referenced in publication

**8 Code and data availability**

The version of FESOM2.0 used to carry out simulations reported here can be accessed from https://swrepo1.awi.de/svn/awi-fvom/ after registration. The updated versions will be available through the same link in future. For convenience, the configuration used, together with the meshes, is archived at doi:10.5281/zenodo.161319. Mesh partitioning in FESOM is based on a METIS Version 5.1.0 package developed at the Department of Computer Science & Engineering at the University of Minnesota (http://glaros.dtc.umn.edu/gkhome/views/metis). METIS and pARMS (Li et al., 2003) present separate libraries which are freely available subject to their licenses. FESOM1.4 is available at https://swrepo1.awi.de/projects/fesom/ (requires registration). The Polar Science Center Hydrographic Climatology (Steele et al., 2001) used to initialize runs of CORE-II atmospheric forcing data (Large and Yeager, 2009) is freely available online. The simulation results can be obtained from the authors on request.

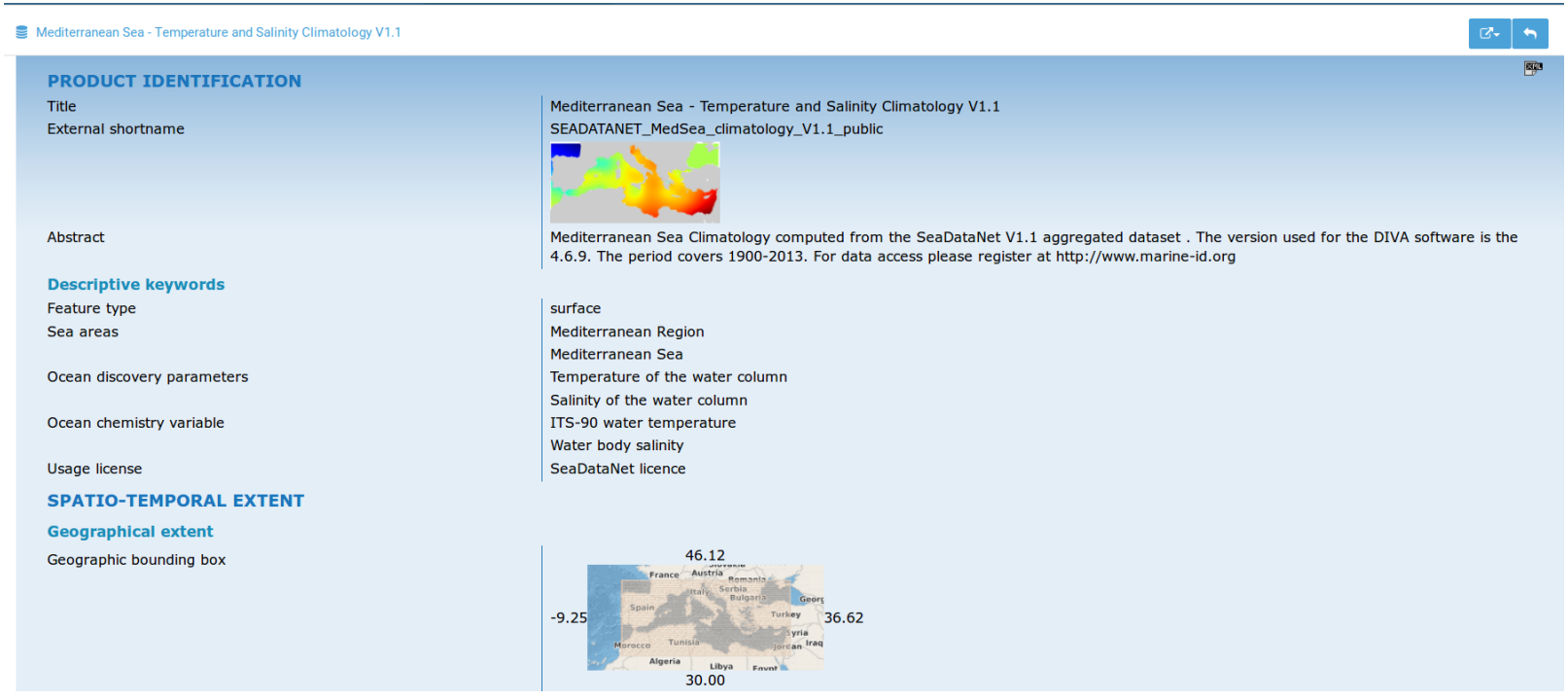
# Products and results

**Results:**
numerical model outputs (re-analysis, forecasts)
climatologies build from in situ data
aggregated datasets

**Goals:**
proper citation in publications
control of different versions of the same product

# Products and results

**Example:** SeaDataNet Product Catalog (Sextant)
Mediterranean Sea: Temperature and Salinity Climatology V1.1

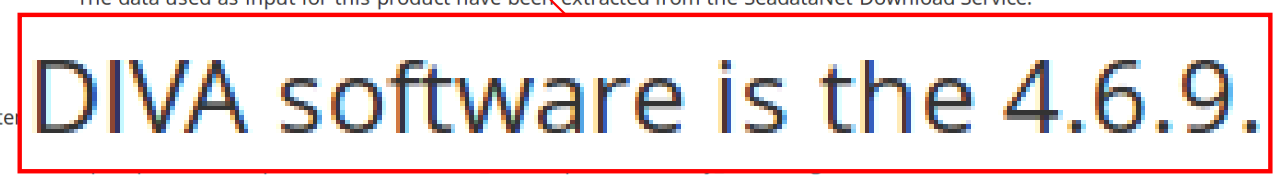



- Internal permanent shortname: 90ae7a06-8b08-4afe-83dd-ca92bc99f5c0
- DOI: 10.12770/90ae7a06-8b08-4afe-83dd-ca92bc99f5c0

# Products and results

## Mediterranean Sea - Temperature and Salinity Climatology V1.1

| | |
|---|---|
| Date(s) | 2015-11-30 (Creation) |
| Custodian(s): | IFREMER / IDM/SISMER |
| Originator(s): | Istituto Nazionale di Geofisica e Vulcanologia – INGV, Sede di Bologna |
| Credit | Seadatanet |
| Version | 1.1 |
| DOI | 10.12770/90ae7a06-8b08-4afe-83dd-ca92bc99f5c0 |
| Abstract | Mediterranean Sea Climatology computed from the SeaDataNet V1.1 aggregated dataset . The version used for the DIVA software is the 4.6.9. The period covers 1900-2013.<br>For data access please register at http://www.marine-id.org |
| Keywords | Oceanographic geographical features, Temperature of the water column, Salinity of the water column, ITS-90 water temperature, Water body salinity, Mediterranean Region, Mediterranean Sea |
| Lineage | The data used as input for this product have been extracted from the SeadataNet Download Service: |
| Utilisation | |
| Temporal Exte | |
| Data | /SDN_2015-11_TS_Med_Sea_Climatology_v1.1.zip |




Can we go further and have:

"*The version used for the DIVA software is the 4.6.9, doi: 10.5281/zenodo.836727*" ?

# Software tools and methods: Journals

## Geoscientific Model Development

- *"description, development, and evaluation of numerical models of the Earth system and its components"*
- *"geoscientific model descriptions, from statistical models to box models to GCMs"*
- *"Inclusion of Code and/or data availability sections is mandatory for all papers"*

Geosci. Model Dev., 7, 225-241, 2014
https://doi.org/10.5194/gmd-7-225-2014


Volume 7, issue 1



Methods for assessment of models

29 Jan 2014

# divand-1.0: *n*-dimensional variational data analysis for ocean observations

**A. Barth[1,*], J.-M. Beckers[1], C. Troupin[2], A. Alvera-Azcárate[1], and L. Vandenbulcke[3,4]**
[1]GHER, University of Liège, Liège, Belgium
[2]IMEDEA, Esporles, Illes Balears, Spain
[3]seamod.ro/Jailoo srl, Sat Valeni, Com. Salatrucu, Jud. Arges, Romania
[4]CIIMAR, University of Porto, Porto, Portugal
[*]Invited contribution by A. Barth, recipient of the EGU Arne Richter Award for Outstanding Young Scientists 2010.

Received: 07 Jun 2013 – Discussion started: 23 Jul 2013
Revised: 18 Oct 2013 – Accepted: 12 Dec 2013 – Published: 29 Jan 2014

doi:10.5194/gmd-7-225-2014

# Software tools and methods: Journals

## Earth Science Informatics

- *"(...) cutting-edge, and provocative scientific work in the area of Earth Science Informatics (...)"*
- *"(...) all aspects of computer applications to the acquisition, storage, processing, interchange, and visualization of data"*
- *Sub-disciplines: Ontology, Simulation and Modeling, Information Systems Applications*


doi:10.1007/s12145-016-0266-2

# Software tools and methods: Journals

## Methods in Oceanography

- *"original research on new methods in all aspects of oceanographic research"*
- *"significant advances in the development of new methods for the interpretation of either existing or future data"*


Methods in Oceanography

Volume 17, December 2016, Pages 50-82


Review

### Potential for an underwater glider component as part of the Global Ocean Observing System

T. Liblik [a, b], J. Karstensen [a], P. Testor [c], P. Alenius [d], D. Hayes [e], S. Ruiz [f], K.J. Heywood [g], S. Pouliquen [h], L. Mortier [c], E. Mauri [i]

⊞ Show more

https://doi.org/10.1016/j.mio.2016.05.001 Get rights and content

doi:10.1016/j.mio.2016.05.001

💣 discontinued as of 2017

# Software tools and methods: Research platforms

Definition: online infrastructures whose objective is to persistently store and archive digital artifacts relevant to research:

- articles
- data
- images
- code
- ...

# Software tools and methods: Research platforms

Figshare: "improve the organization of users' research"

- file upload, collaborative spaces, DOI attribution, ...
- on the platform: figures, datasets, media (including video), papers (including pre-prints), posters, code, and filesets.


https://figshare.com/articles/CNR-ISMAR_in_situ_observations_network/4001448

# Software tools and methods: Research platforms

DSpace

- has to be installed on a server manager by the search institution
- Sandbox: `http://demo.dspace.org/`

According to the Registry of Open Access Repositories (ROAR)

| | |
|---|---|
| 1759 | institutions or companies as users |
| 1374 | repositories dedicated to Institutional or Departmental Research |
| 32 | repositories dedicated to Research Data |
| 3 | dedicated to Open and Linked Data |

# Software tools and methods: Research platforms

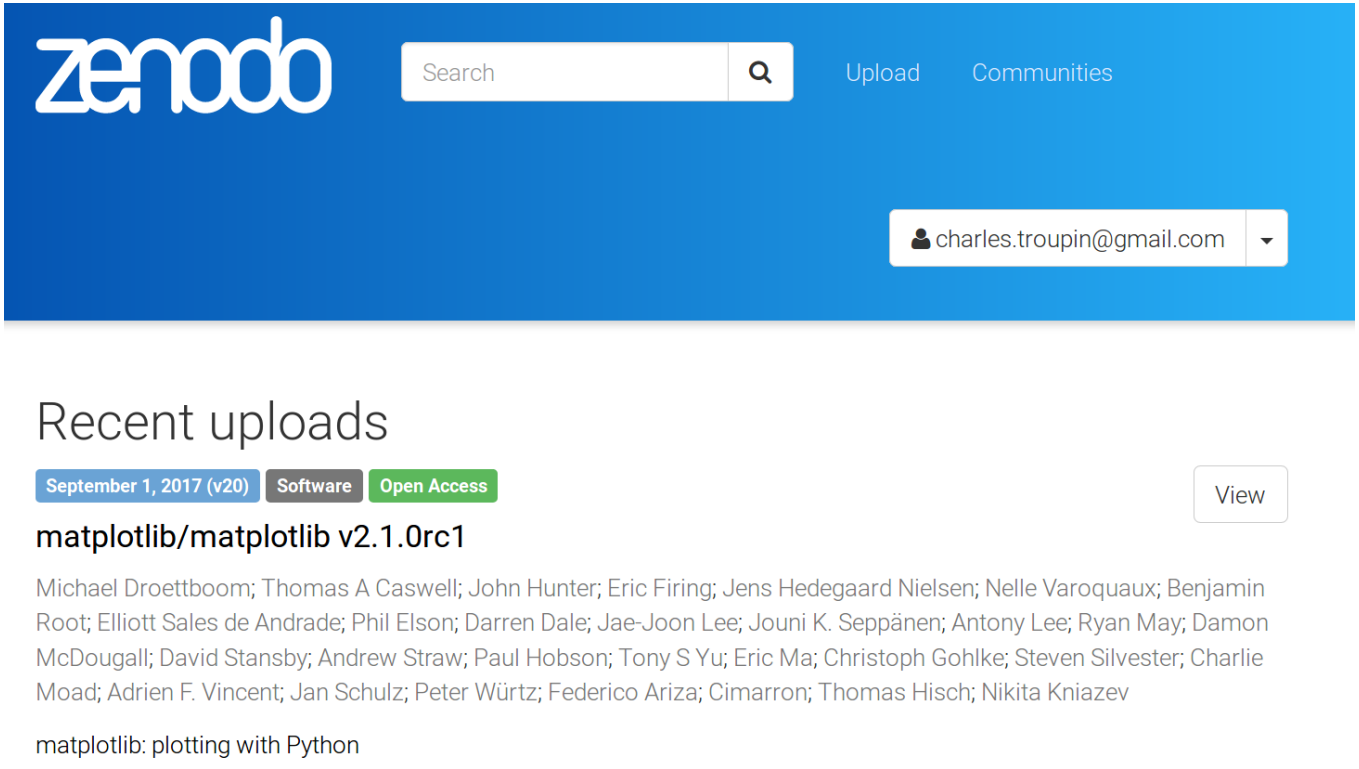
Comprehensive Knowledge Archive Network (CKAN):
can be considered as a tool for making open data websites

- streamline publishing, sharing, finding and using data
- used by numerous governments, organisations and communities around the world
- has to be installed on a server manager by the search institution

# Software tools and methods: Research platforms

Zenodo: research data repository
funded by the European Commission

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



# Comparison

| Tool | CKAN | DSpace | Figshare | Zenodo |
|---|---|---|---|---|
| Open Source | Yes | Yes | No | Yes |
| | ckan/ckan | DSpace/DSpace | – | zenodo/zenodo |
| Licence | Affero GNU GPL v3.0 | BSD | – | GPL-2.0 |
| 1st released | November 2011 | November 2002 | January 2011 | May 2013 |
| Main technology | Python | Java | – | Python |
| Deployment | Local | Local | Cloud | Cloud |
| Integration with | No | No | Yes | Yes |
| Integration with | Yes | Not direct | Yes | Login |

# Comparison

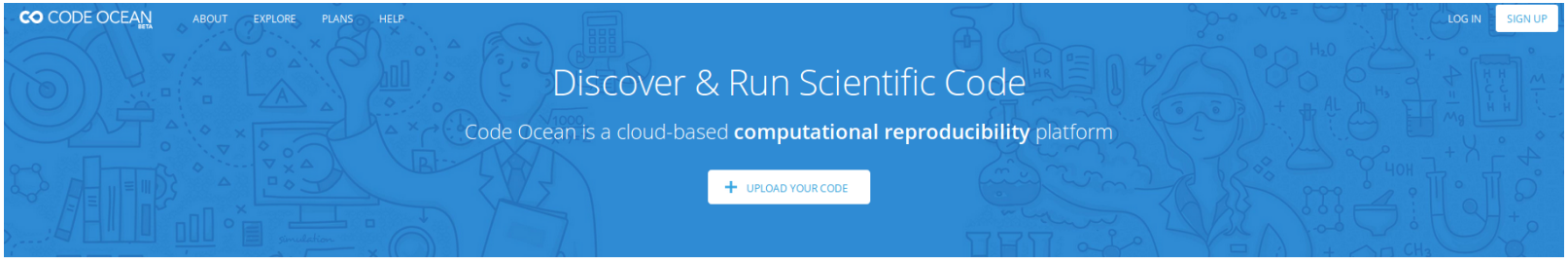
| Tool | CKAN | DSpace | Figshare | Zenodo |
|---|---|---|---|---|
| Open Source | Yes | Yes | No | Yes |
| | ckan/ckan | DSpace/DSpace | – | zenodo/zenodo |
| Licence | Affero GNU GPL v3.0 | BSD | – | GPL-2.0 |
| 1st released | November 2011 | November 2002 | January 2011 | May 2013 |
| Main technology | Python | Java | – | Python |
| Deployment | Local | Local | Cloud | Cloud |
| Integration with | No | No | Yes | Yes |
| Integration with | Yes | Not direct | Yes | Login |

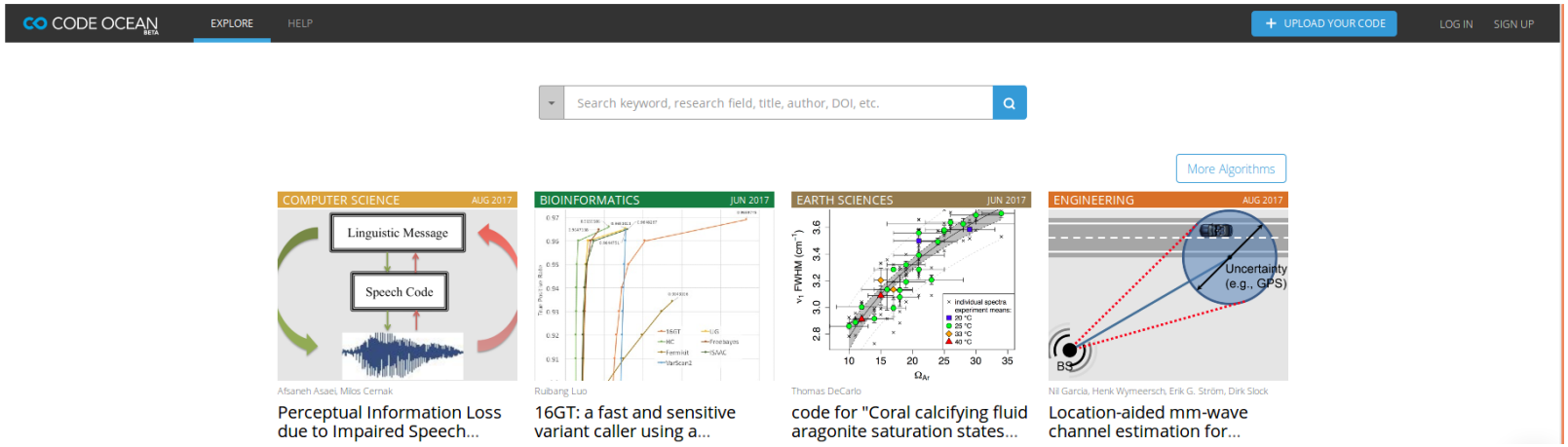
## Choice for further tests: Zenodo

1. Free and open software
2. Cloud service, i.e. no installation
3. Coupling with GitHub
4. Login via ORCID

# Computational reproducibility platform

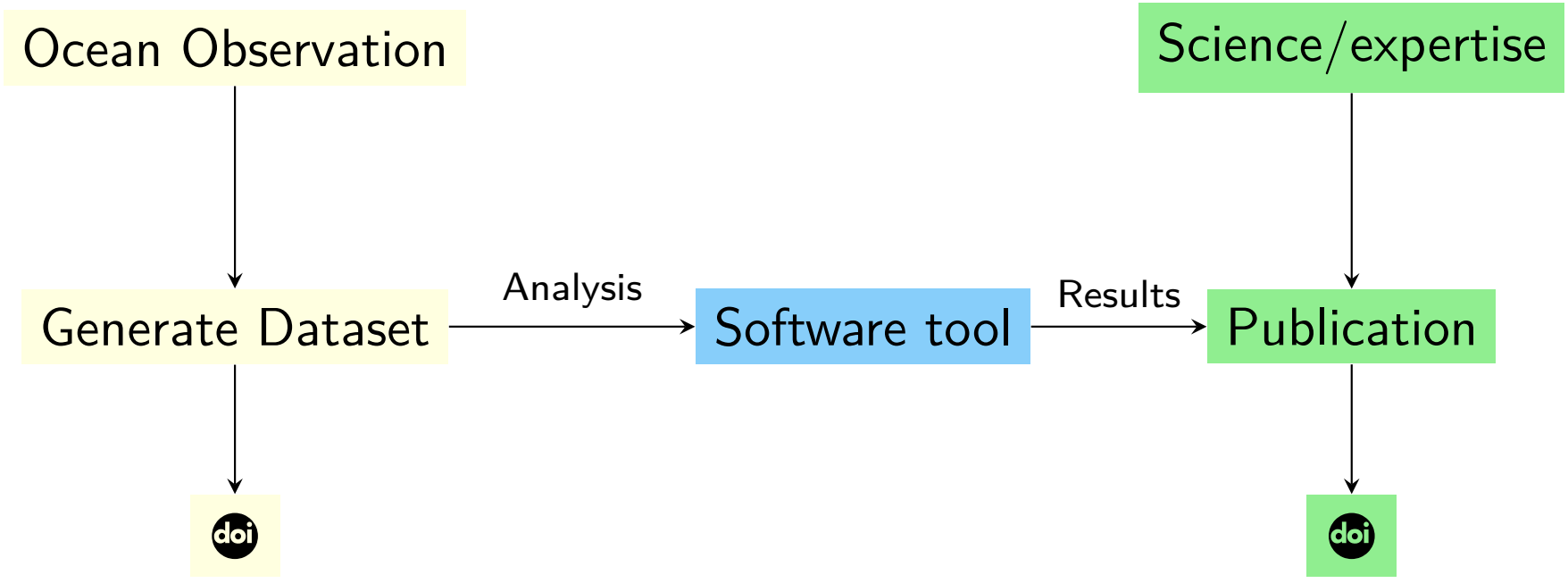
Code Ocean: "easy way to share, discover and run code published in academic journals and conferences"

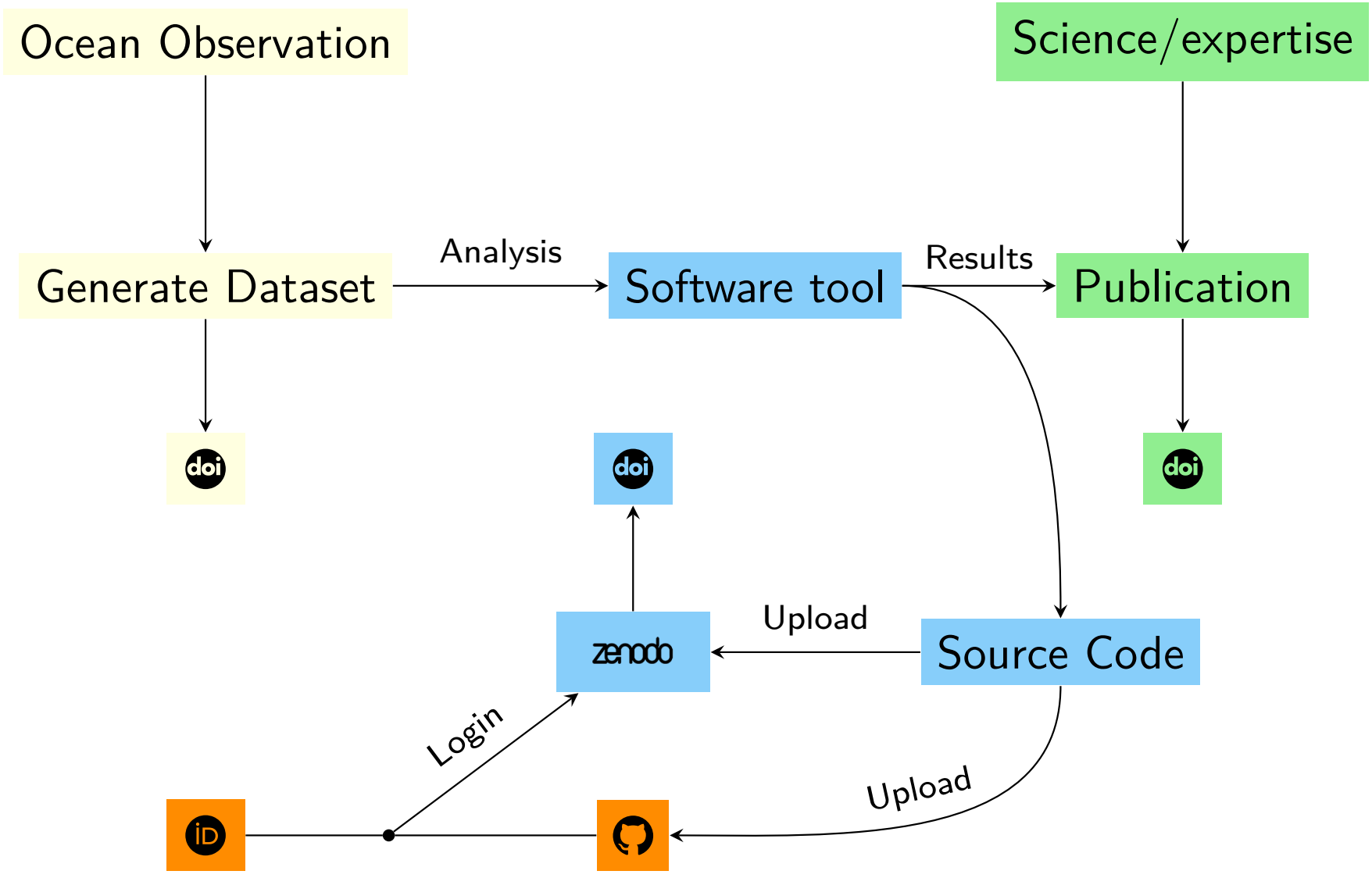


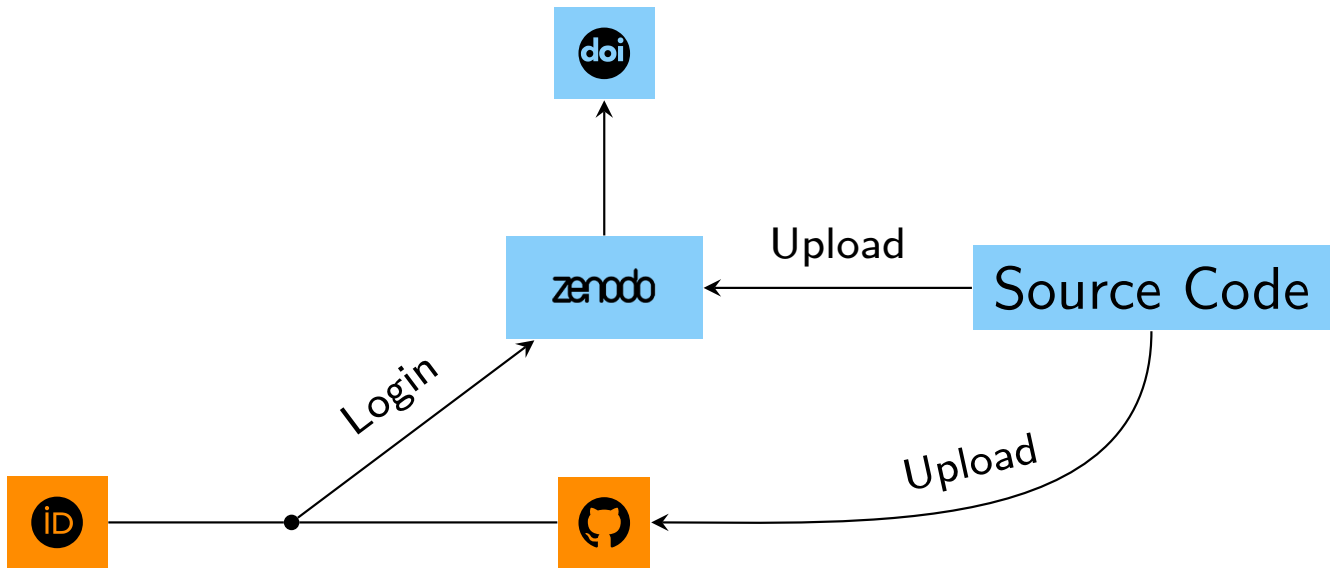



# *Motivations:*

Reproducibility & Traceability


Ocean Observation
Science/expertise
Generate Dataset
Analysis
Software tool
Results
Publication
doi
doi


Ocean Observation
Science/expertise
Generate Dataset
Analysis
Software tool
Results
Publication
doi
doi
doi
zenodo
Upload
Source Code
Login
iD
Upload


doi
zenodo
Upload
Source Code
Login
iD
Upload

# Goals

Reproducibility: **IF** same experiment
identical parameters
same dataset
same model
**THEN** same results

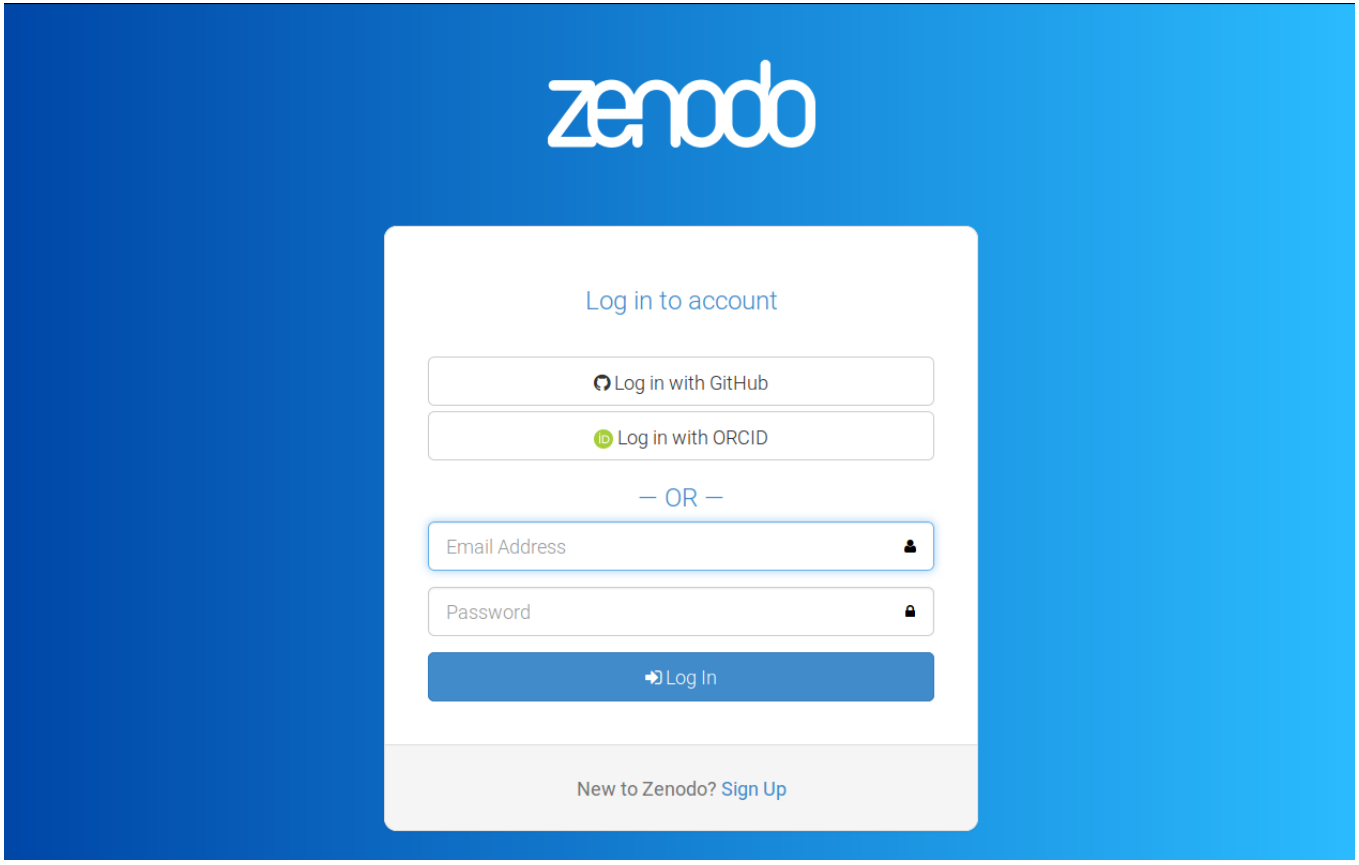
Traceability: all the elements used in the analysis/experiment:
accessible
properly described
uniquely identified

# *A closer look*

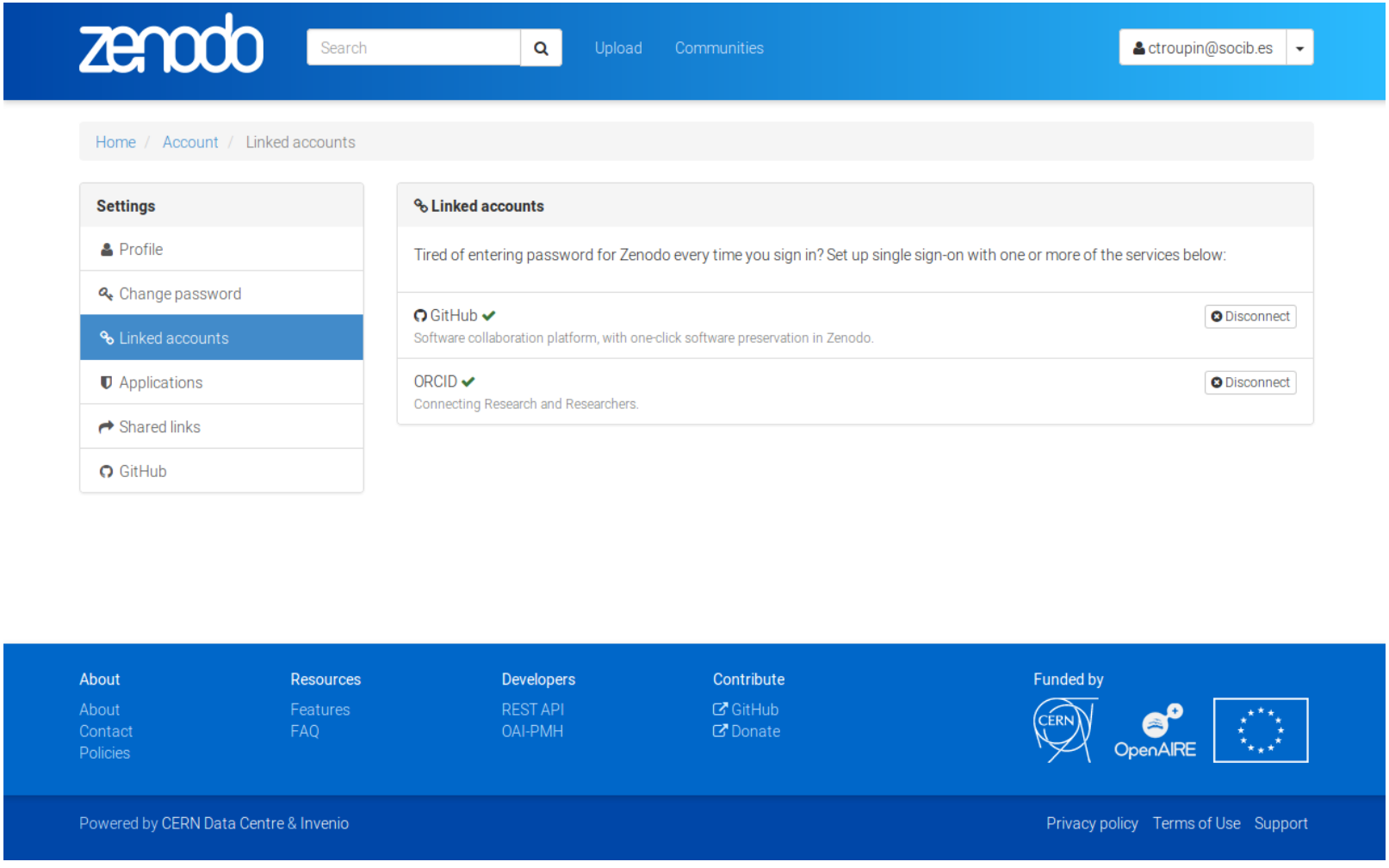
to Zenodo

# Login: 3 options

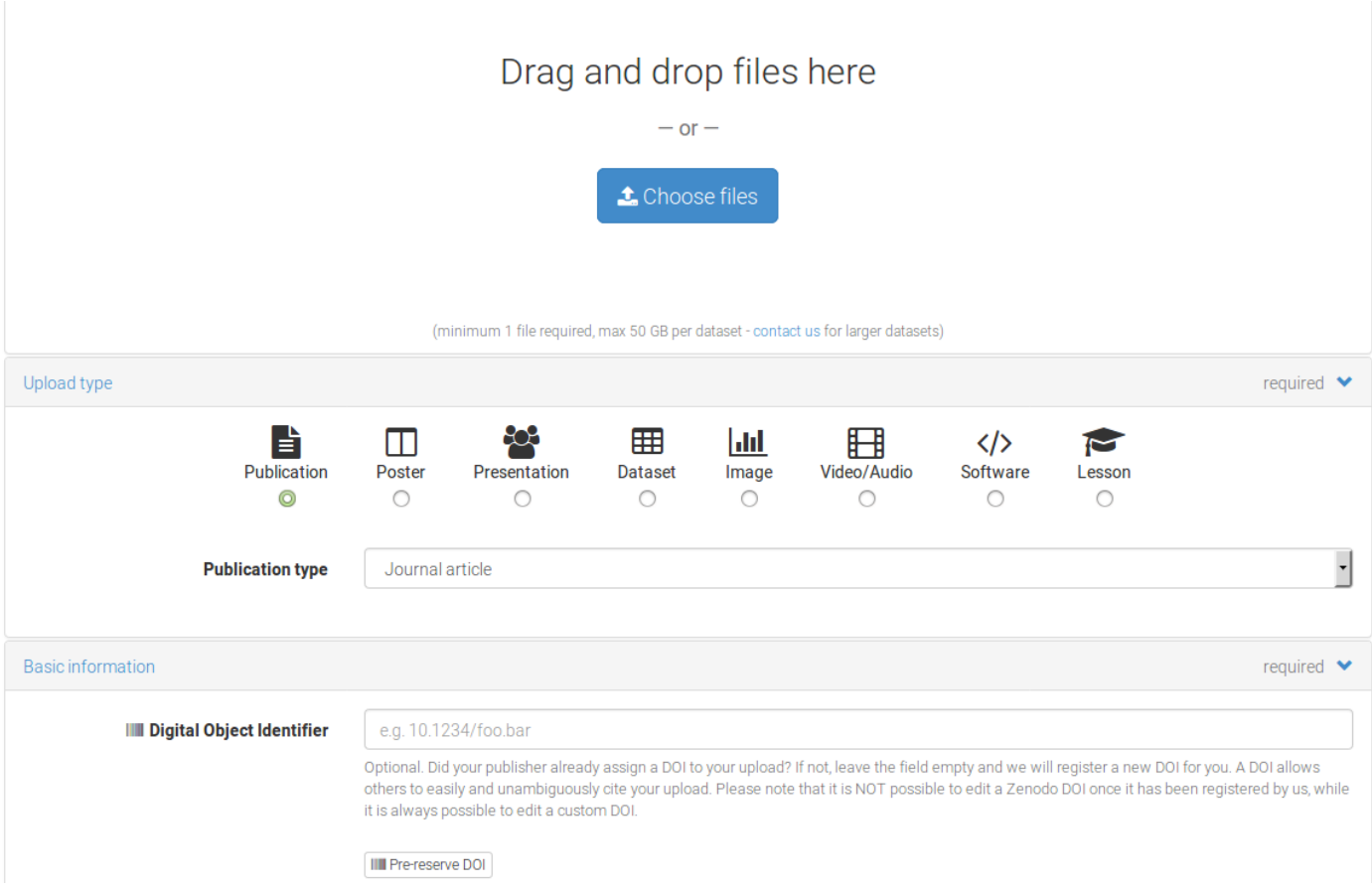
1. Use  account
2. Use  account
3. Create new  (Sign Up)



# Main page: linked accounts



# Upload: you can drop anything




In particular: Software

(ok for stable code)

# Upload: add metadata

**Digital Object Identifier** 10.5281/zenodo.836727

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.

Reserve DOI ✔

**Publication date** * 2017-07-31

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

**Title** * gher-ulg/DIVA: v4.7.1

Required.

**Authors** *

| | |
|---|---|
| Sylvain Watelet | GHER, University of Liège |
| Charles Troupin | GHER, University of Liège |
| Jean-Marie Beckers | GHER, University of Liège |
| Alexander Barth | GHER, University of Liège |
| Mohamed Ouberdous | GHER, University of Liège |

# Upload: add metadata

**Digital Object Identifier** 10.5281/zenodo.836727

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.

Reserve DOI ✔

**Publication date** * 2017-07-31

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

**Title** * gher-ulg/DIVA: v4.7.1

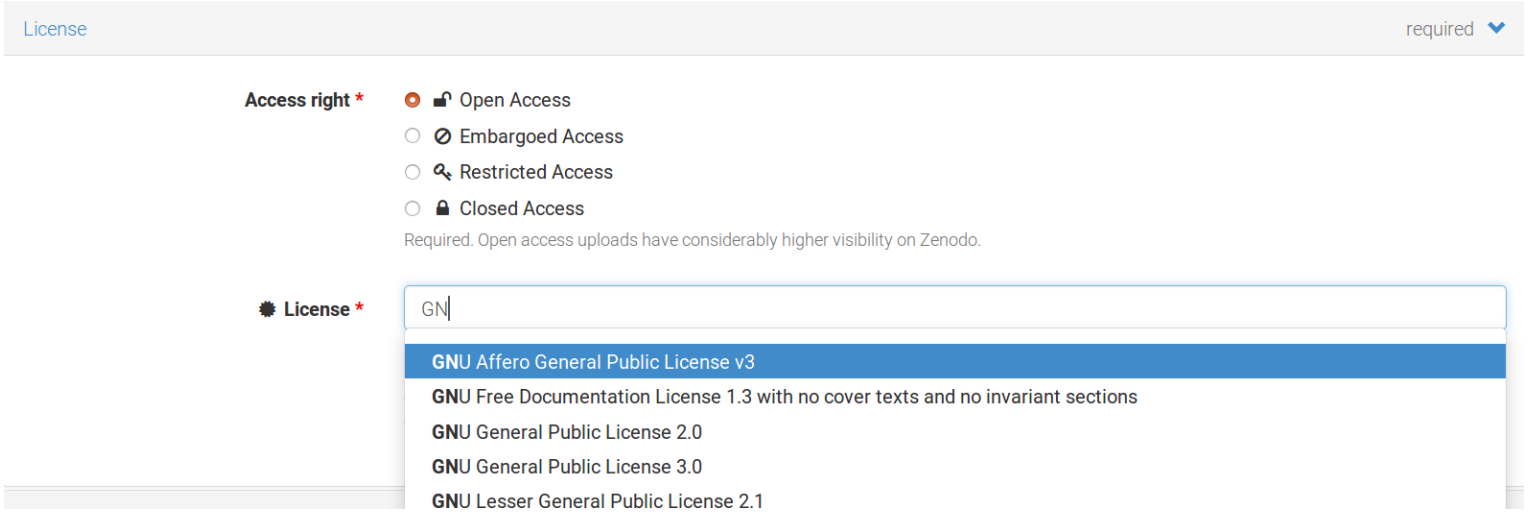
Required.

**Authors** *

| | | |
|---|---|---|
| Sylvain Watelet | GHER, University of Liège | 0000-0002-0265-1021 |
| Charles Troupin | GHER, University of Liège | ORCID (eg. 0000-0002-1825-0097) |
| Jean-Marie Beckers | GHER, University of Liège | ORCID (eg. 0000-0002-1825-0097) |
| Alexander Barth | GHER, University of Liège | ORCID (eg. 0000-0002-1825-0097) |
| Mohamed Ouberdous | GHER, University of Liège | ORCID (eg. 0000-0002-1825-0097) |

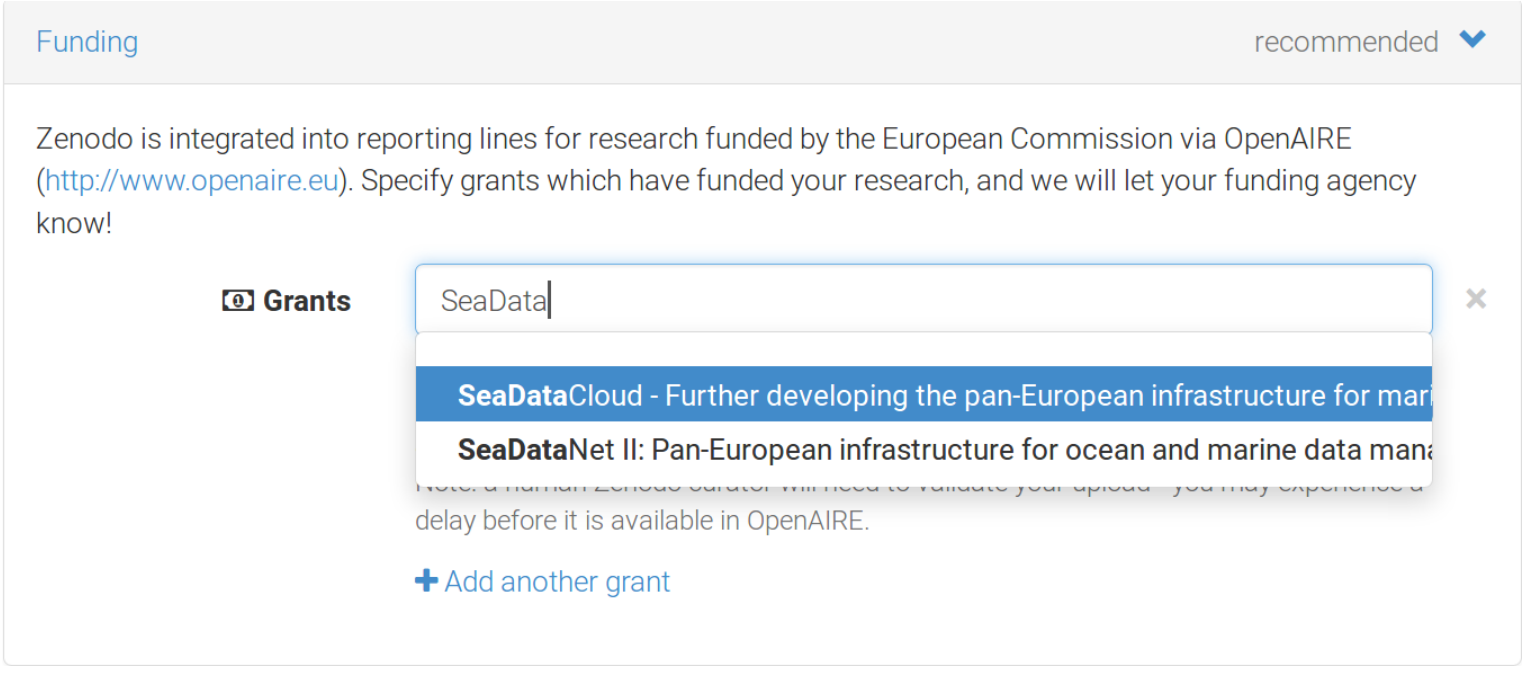
Optional.

# Upload metadata: licence



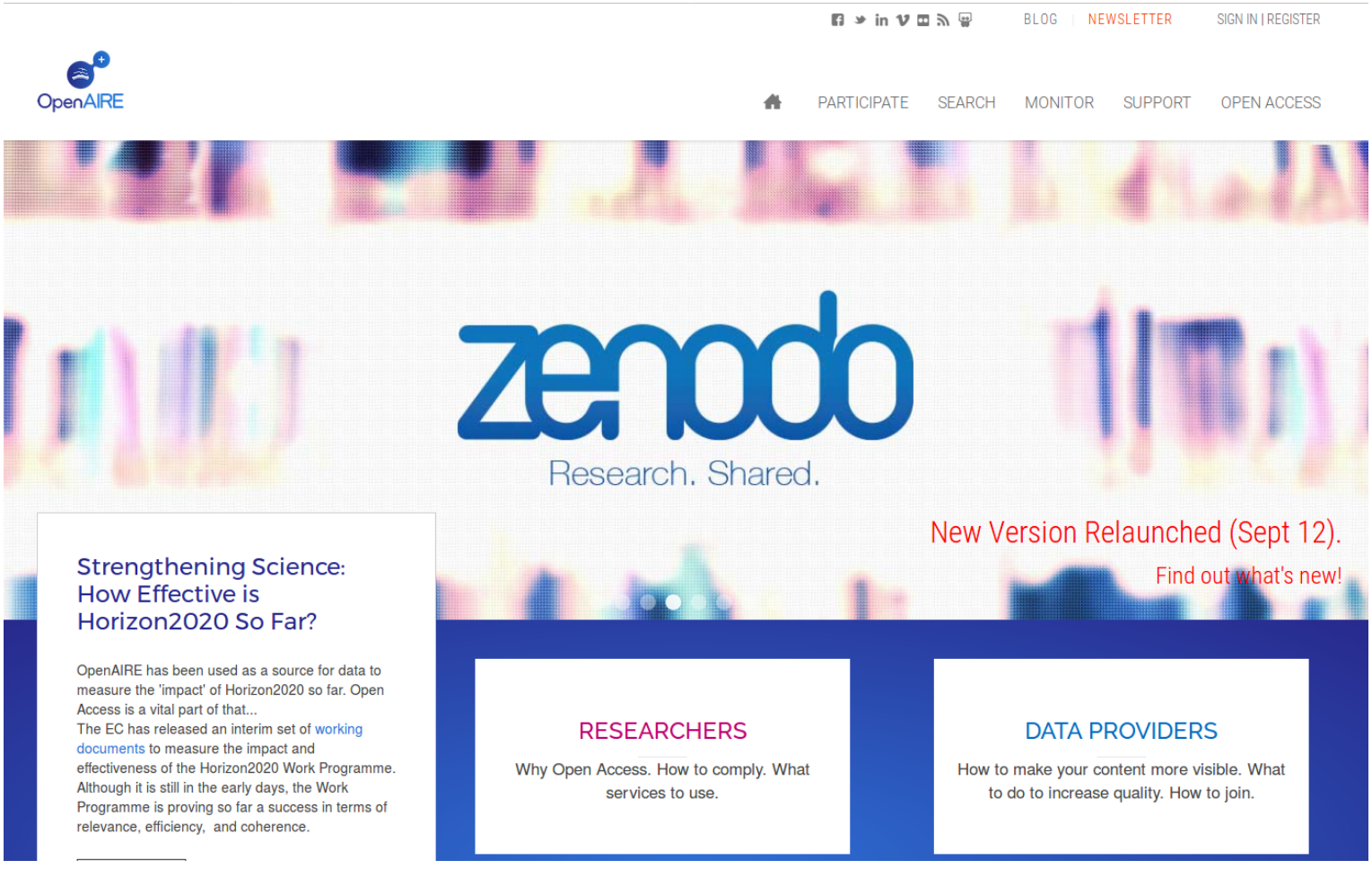

Not necessarily Open!

# Upload metadata: funding!



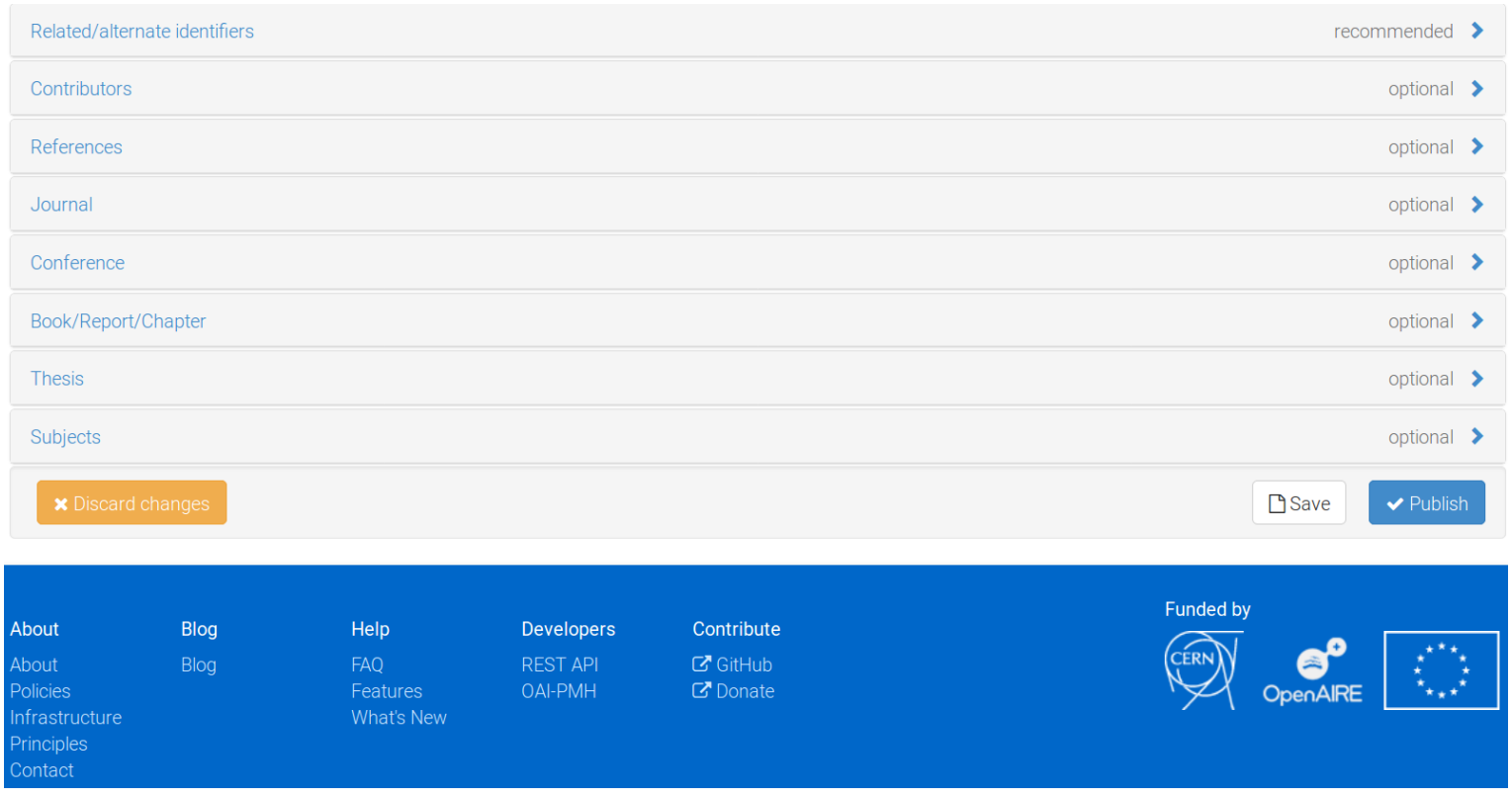

Common strategy for SeaDataCloud, ODIP, ...tools?

# Upload metadata: funding!




Common strategy for SeaDataCloud, ODIP, ...tools?

# Upload: finish

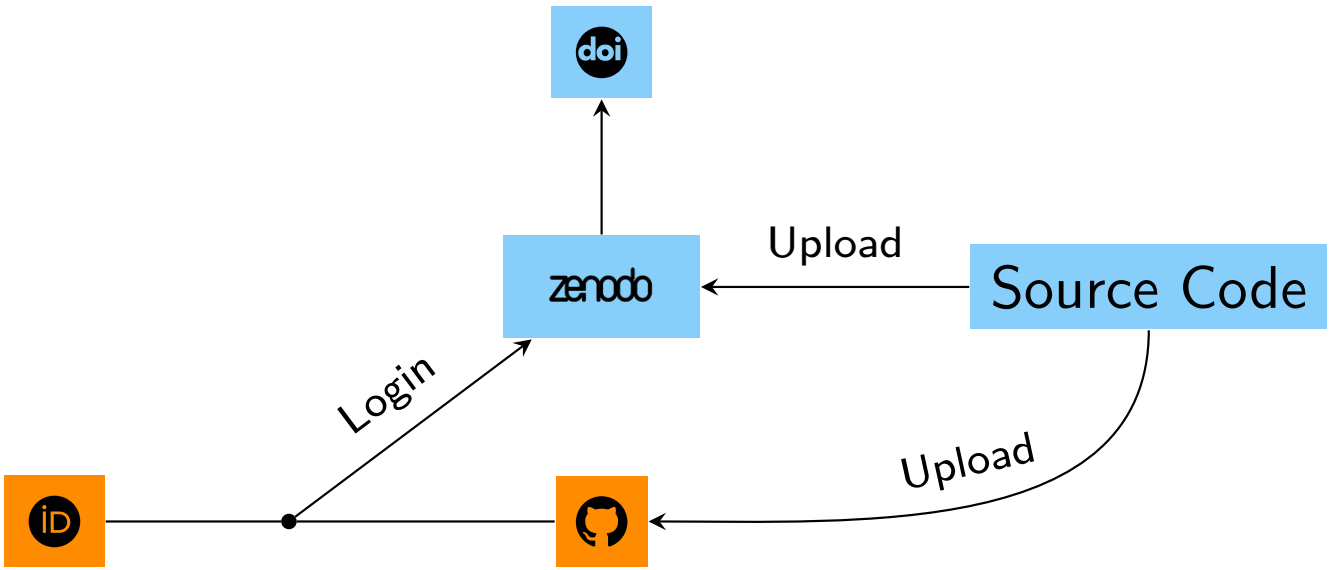


# *Zenodo & Github*

working hand in hand

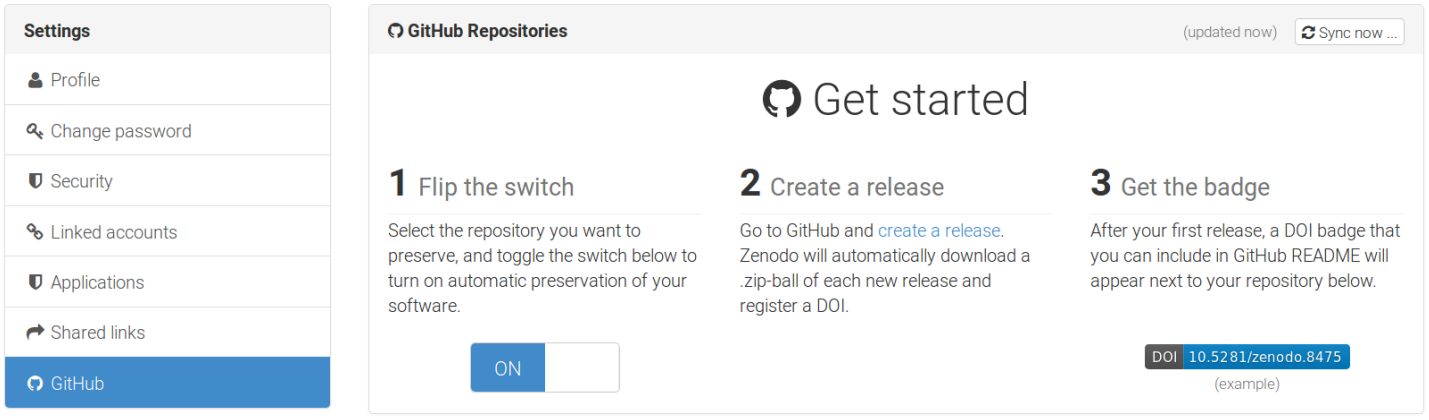
# Requirements

1. or account
2. Source code



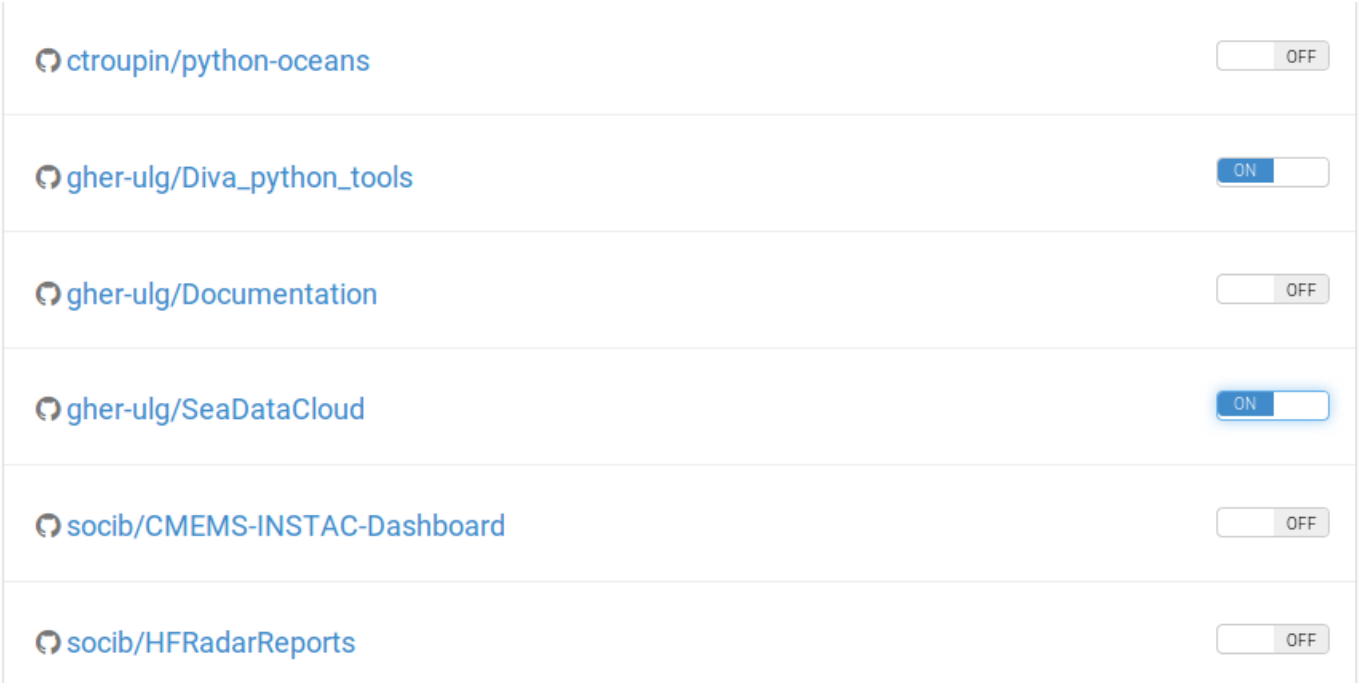
# Generating doi for software releases

In Zenodo: click on the GitHub tab



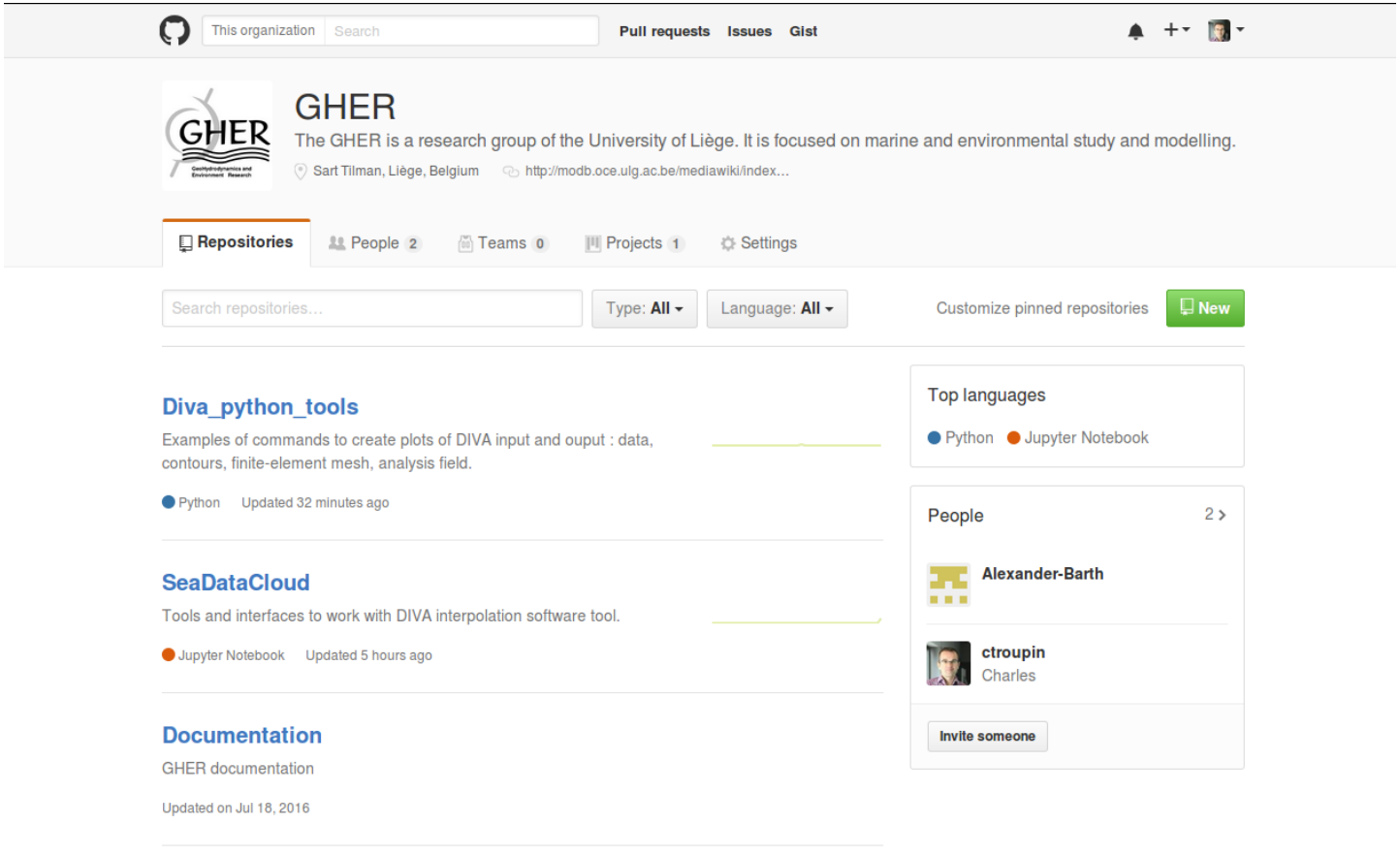
# Generating doi for software releases

In Zenodo: turn on the synchronisation for the selected repositories



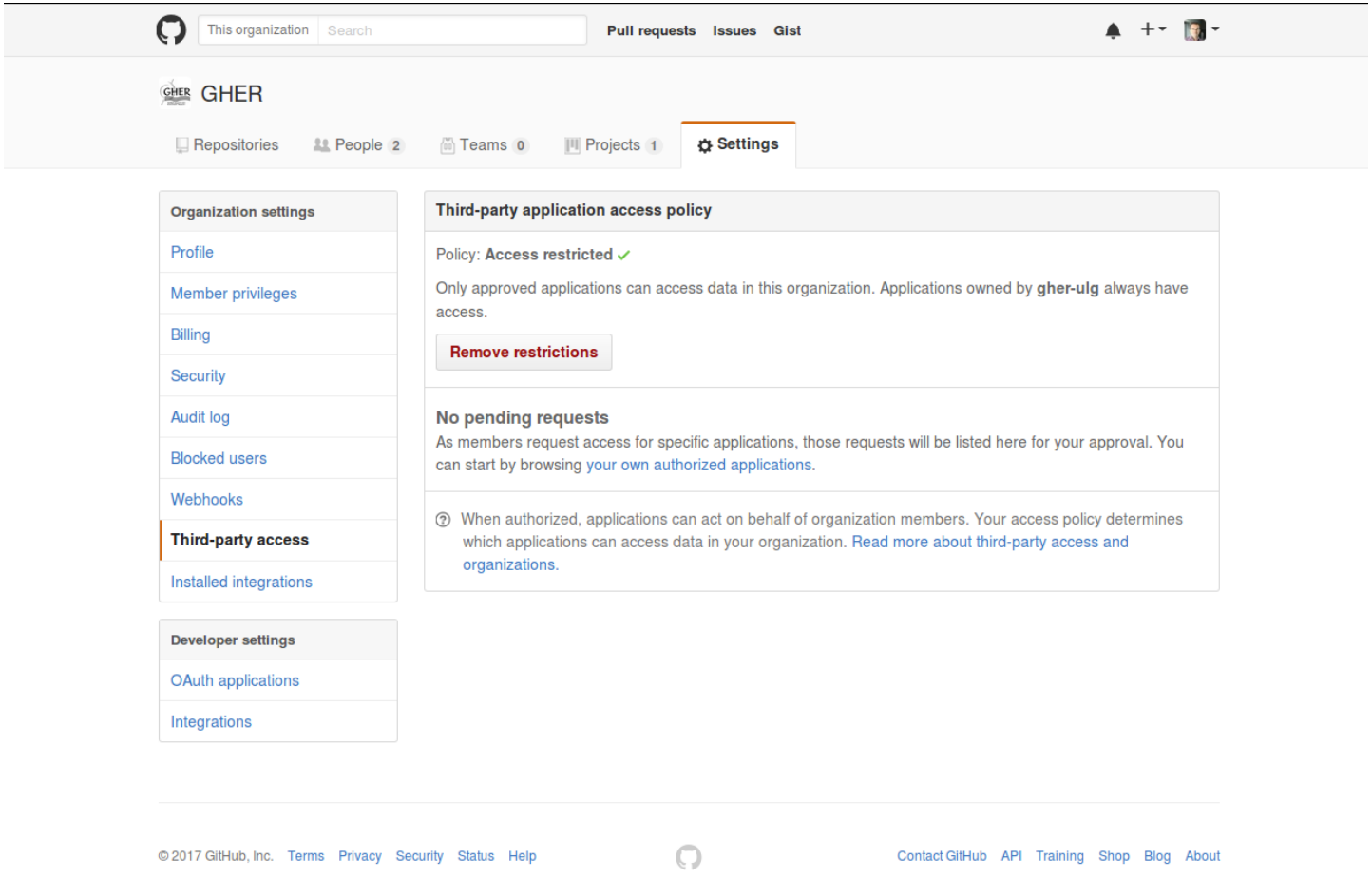
# Generating doi for software releases

## Go on your home page



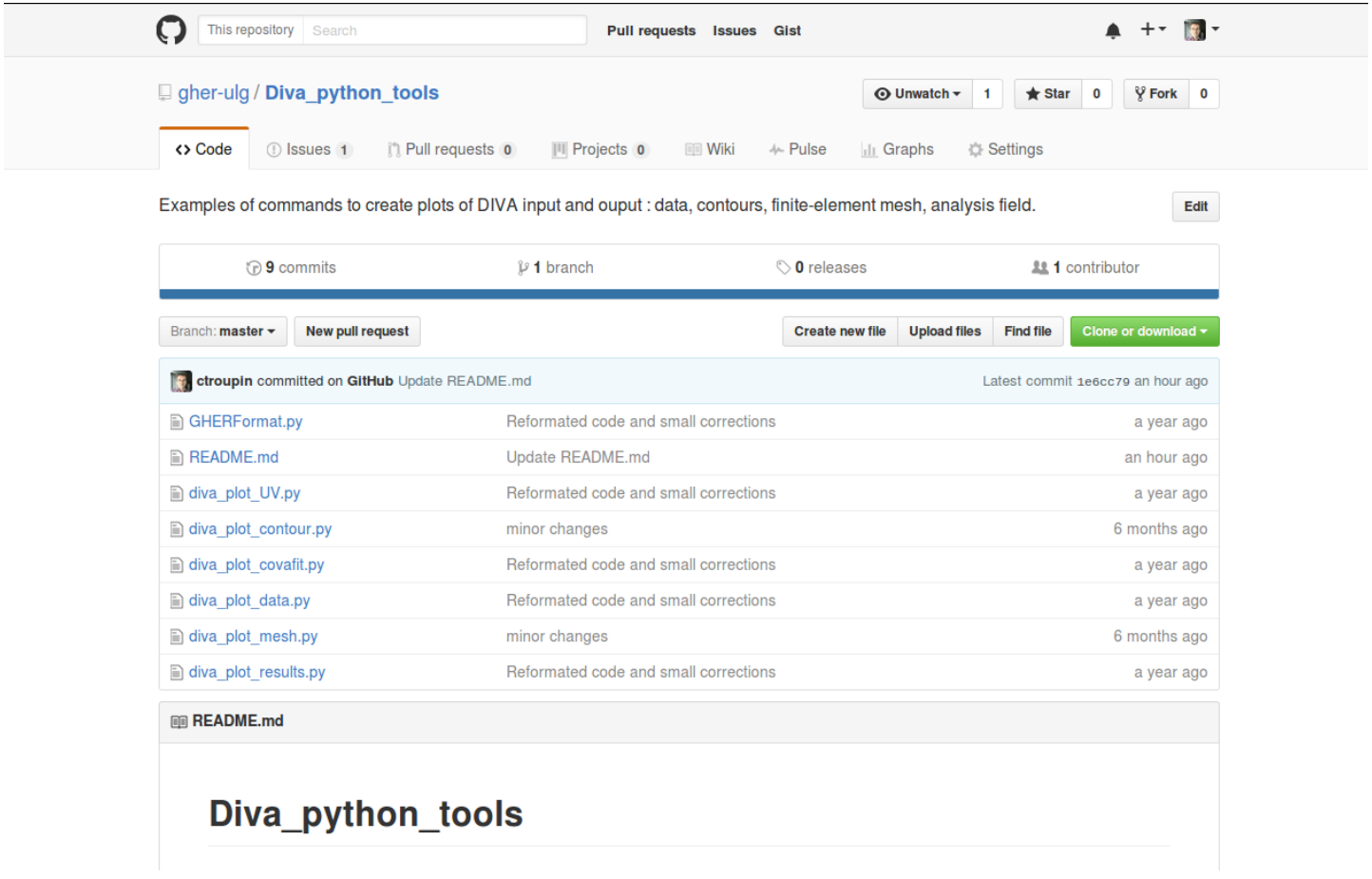
# Generating doi for software releases

## In settings: allow third-party access



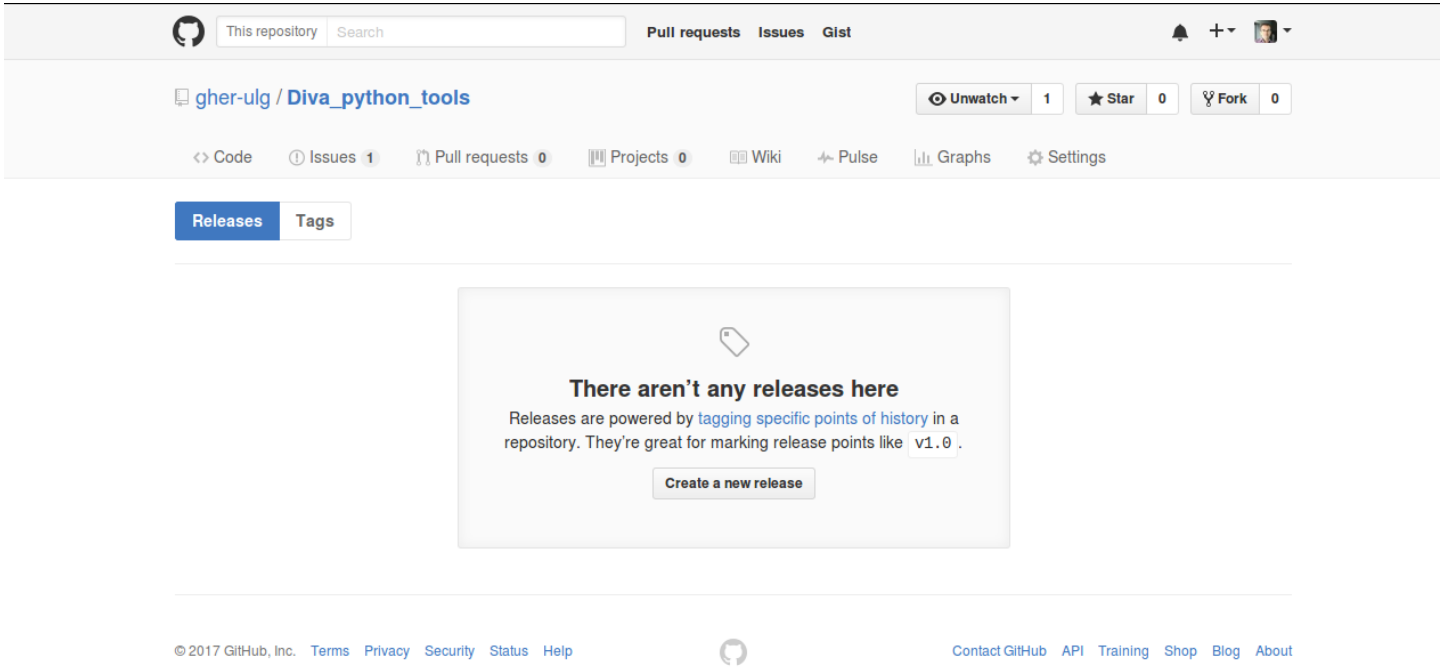
# Generating doi for software releases

## Open the selected project repository



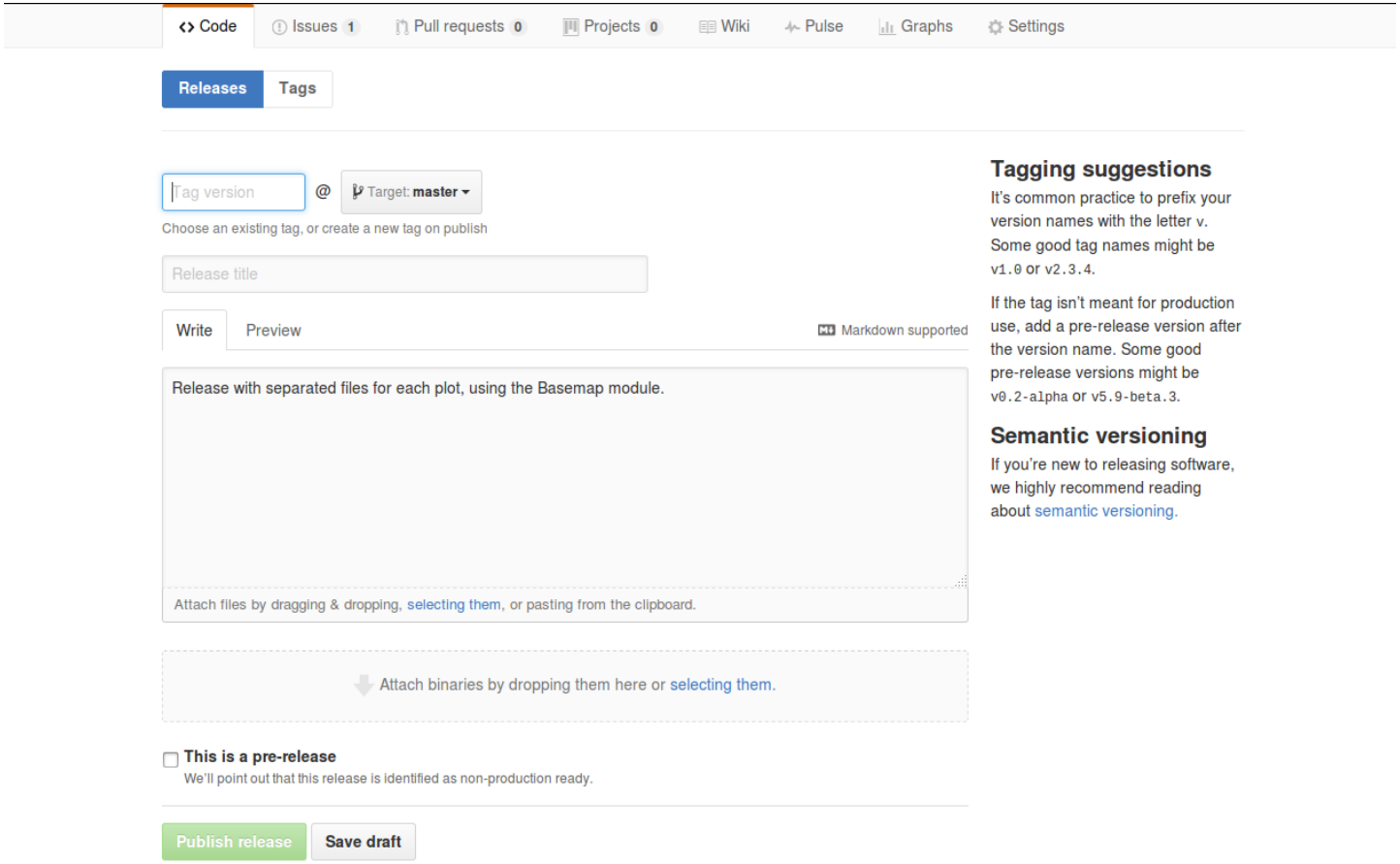
# Generating doi for software releases

Click on the *Release* button



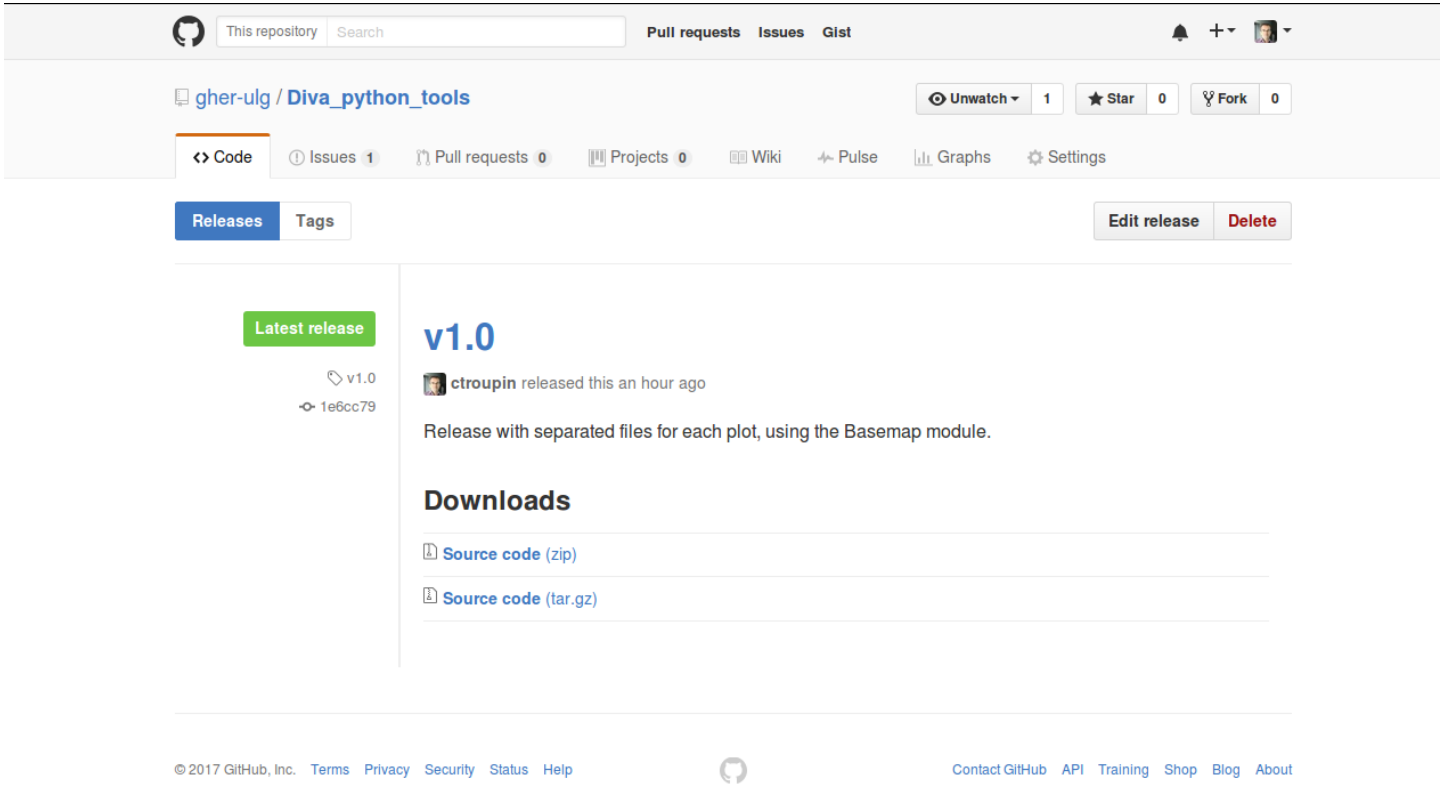
# Generating doi for software releases

Fill in the information and ...



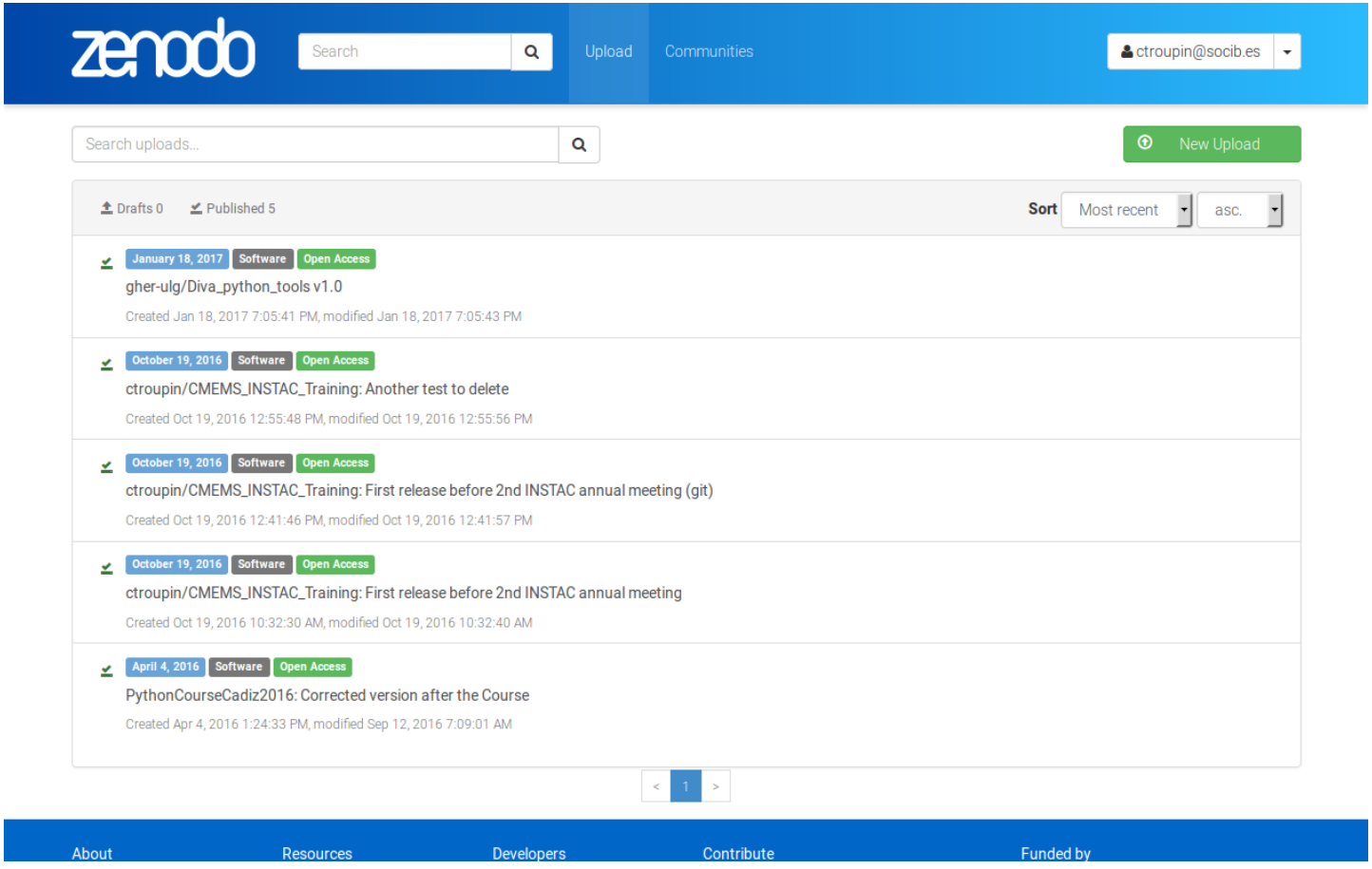
# Generating doi for software releases

...make the release



# Generating doi for software releases

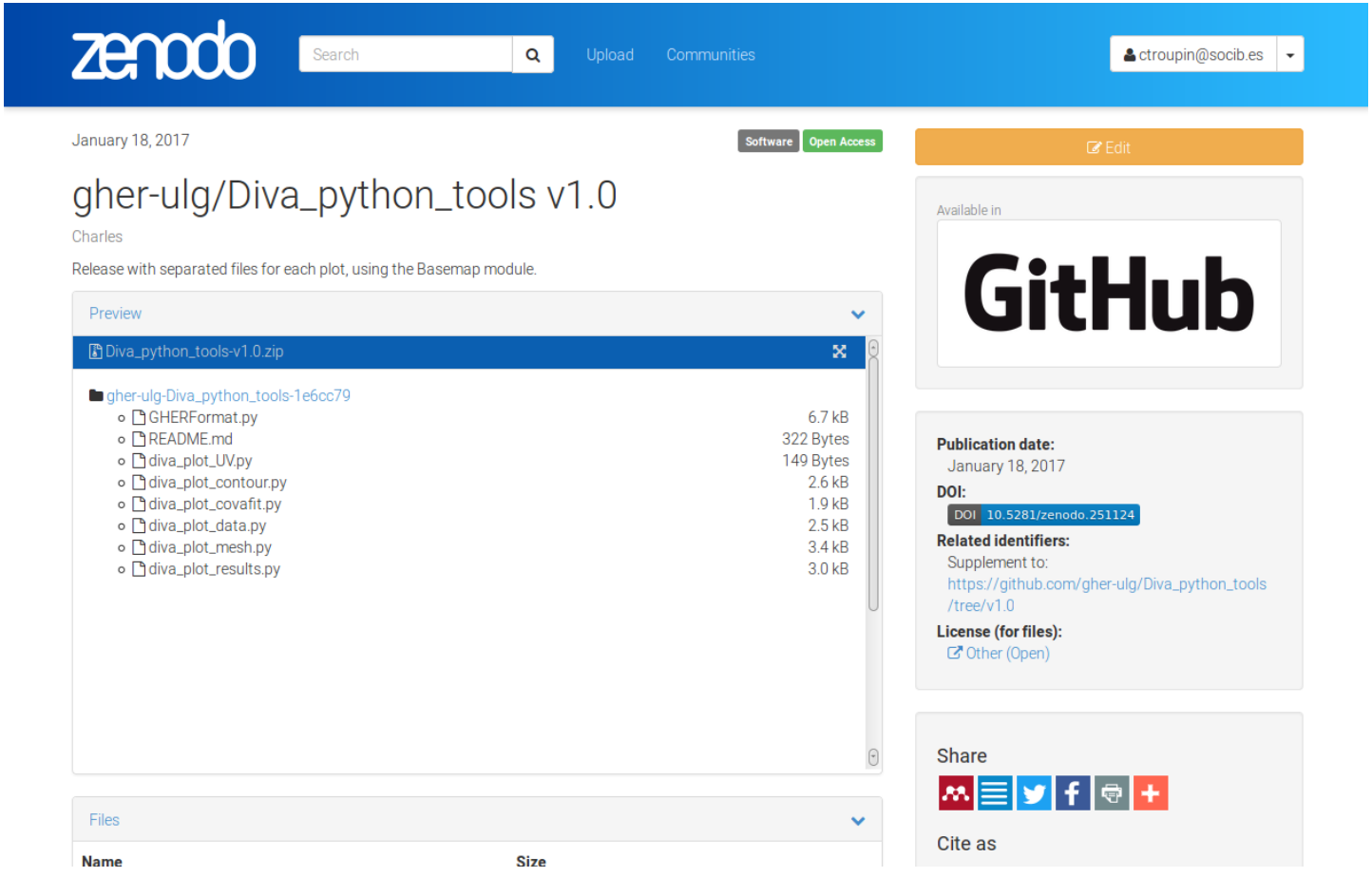
Check the project release on Zenodo and ...



# Generating doi for software releases

...get the doi badge

and celebrate



# *Use case I*

get DOI for Diva releases

# Diva (simplified) development history

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 switch from to **git**, distribution via

# From SVN to Git and GitHub

**Easy way**

1. Create a new repository with the latest release of the code
2. Lose all the history of the changes, the previous releases and developing branches

# From SVN to Git and GitHub

**Hard/conservative way**

1. Git repository whose structure reflects that of SVN (trunk, branches, tags)
2. Use the `git-svn` bridge
3. End up with a new GitHub repos with all the history

# Use-case: Diva releases

1. Switch from SVN to GitHub (conserving the history)

Resources:

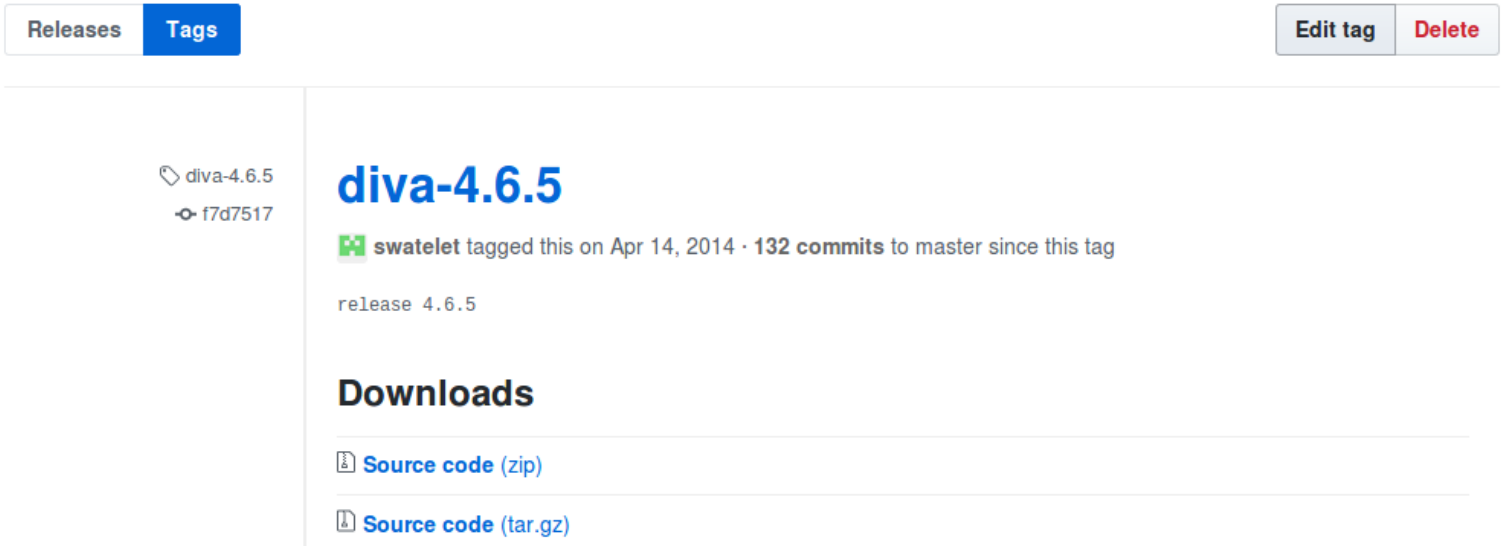
- https://git-scm.com/book/en/v2/Git-and-Other-Systems-Migrating-to-Git
- http://john.albin.net/git/convert-subversion-to-git
- https://www.atlassian.com/git/tutorials/migrating-overview

# Use-case: Diva releases

1. Switch from SVN to GitHub (conserving the history)
2. Enable Diva repository on Zenodo

# Use-case: Diva releases

1. Switch from SVN to GitHub (conserving the history)
2. Enable Diva repository on Zenodo
3. Edit the different *tags* on GitHub to get DOI

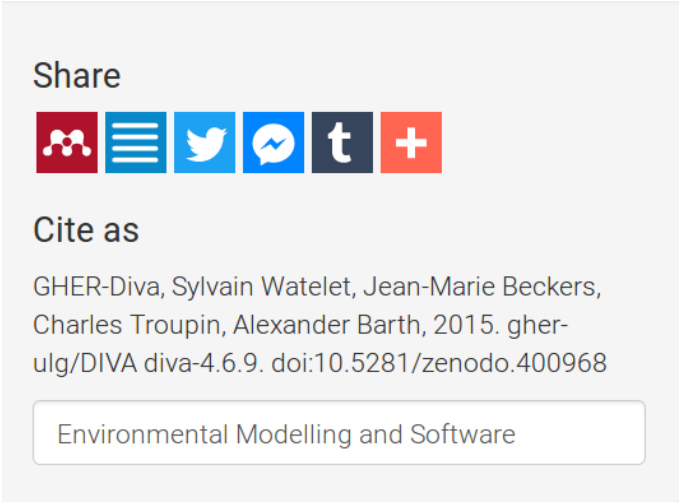


# Use-case: Diva releases

1. Switch from SVN to GitHub (conserving the history)
2. Enable Diva repository on Zenodo
3. Edit the different *tags* on GitHub to get DOI
4. Now we can have
   "*The version used for the DIVA software is the 4.6.9,*
   *doi: 10.5281/zenodo.400968*

# Use-case: Diva releases

1. Switch from SVN to GitHub (conserving the history)
2. Enable Diva repository on Zenodo
3. Edit the different *tags* on GitHub to get DOI
4. Now we can have
   "*The version used for the DIVA software is the 4.6.9,*
   *doi: 10.5281/zenodo.400968*
5. Bonus: "*cite as*" and social media



# *Use case II*

SOCIB Glider toolbox

# Glider toolbox (Matlab/Octave)

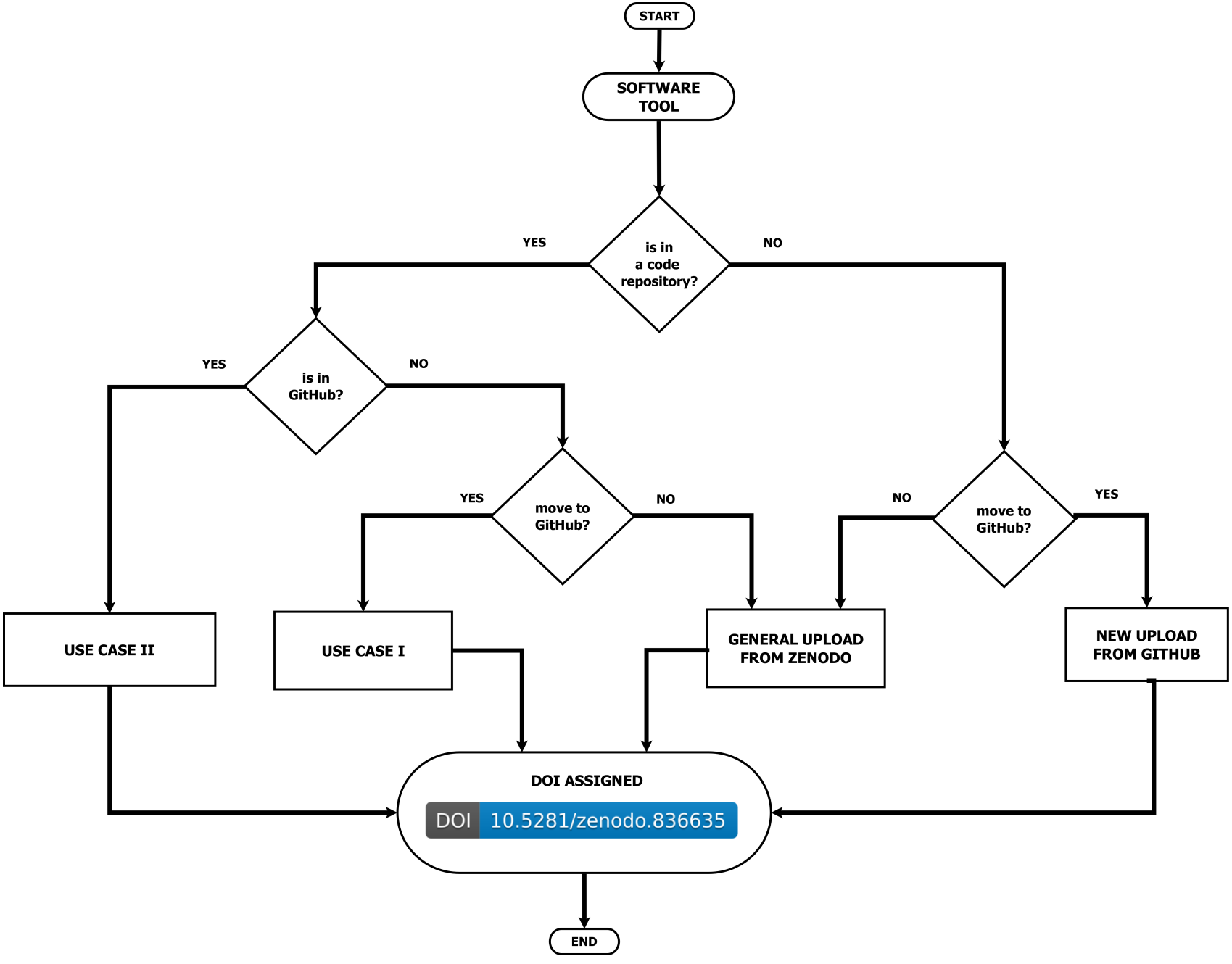
SOCIB Glider toolbox is a set of MATLAB/Octave scripts and functions to manage the data collected by our Glider fleet

↓

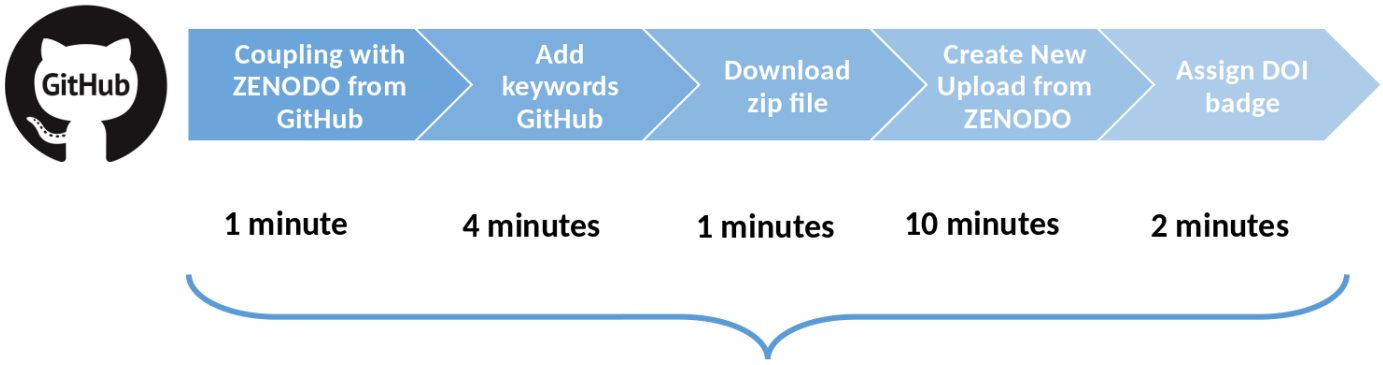
NetCDF-CF compliant

# Motivations

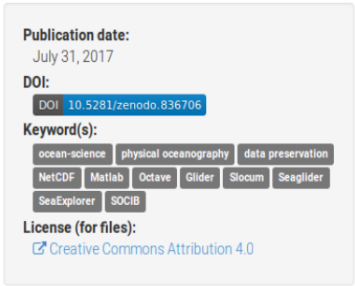
- Publish results of our day by day work
- Track the utilization of our toolbox from research activities


START
SOFTWARE TOOL
is in a code repository?
YES
NO
is in GitHub?
YES
NO
move to GitHub?
YES
NO
move to GitHub?
NO
YES
USE CASE II
USE CASE I
GENERAL UPLOAD FROM ZENODO
NEW UPLOAD FROM GITHUB
DOI ASSIGNED
DOI 10.5281/zenodo.836635
END

# User Experience




**18 minutes**



# Conclusions on Zenodo

1. Seamless integration with GitHub, login via ORCID
2. A big step toward reproducibility
3. A mature and user-friendly tool


*I don't mind your thinking slowly;*
*I mind your publishing faster than you think*

W. Pauli