C. Troupin

GHER-University of Liège

Notebooks, reproducibility and other topics in ocean sciences

The material (slides, exercises) are made available through GitHub at
https://github.com/gher-ulg/COST-EUMETSAT-Training

Who knows/uses ...?

GitHub

Who knows/uses ...?

GitHub

orcID

Who knows/uses ...?

GitHub

orcID

Who knows/uses ...?

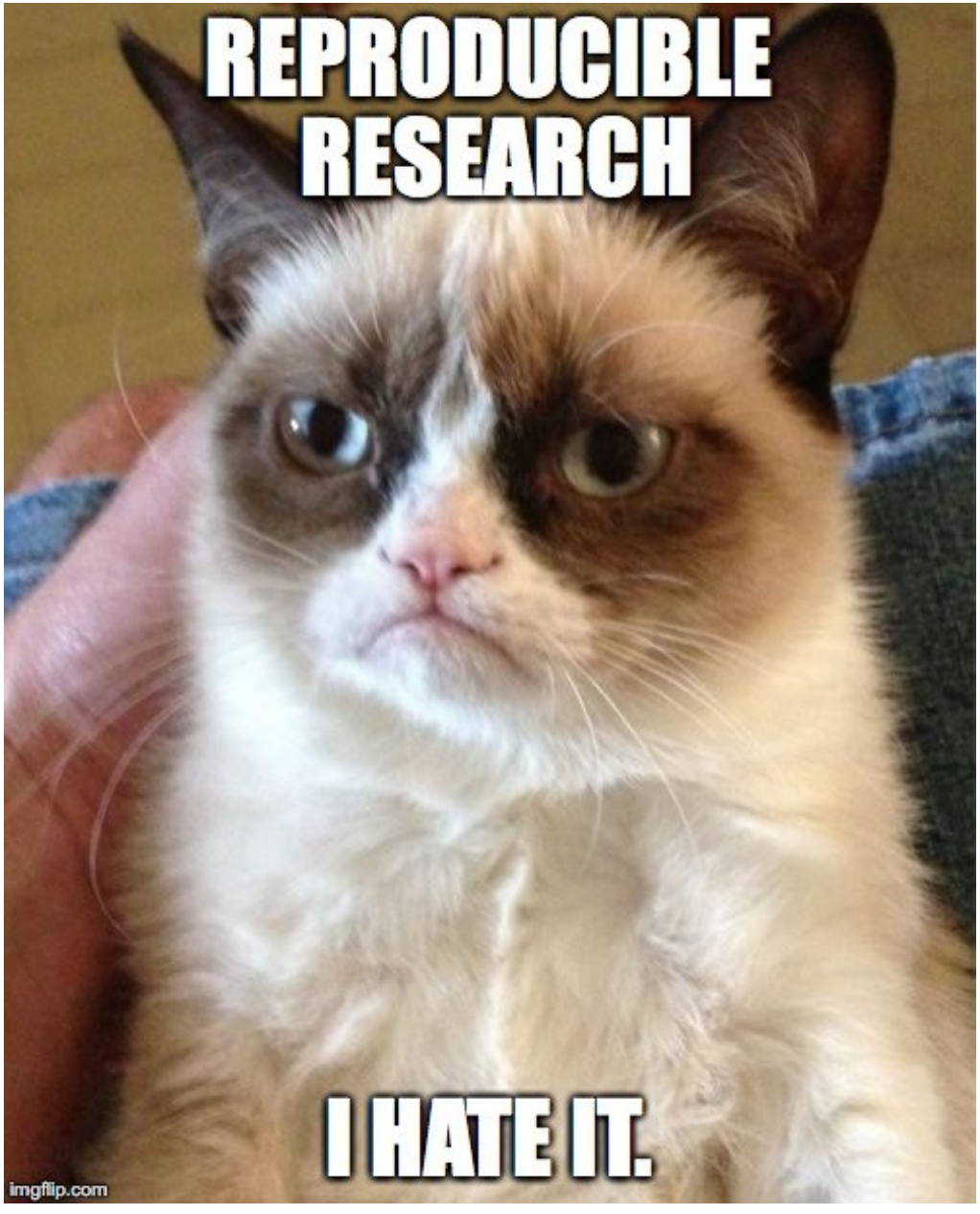
GitHub

orcID

Who knows/uses ...?

GitHub

orcID


REPRODUCIBLE
RESEARCH
I HATE IT.
imgflip.com

Problem:

how to guarantee reproducibility of results?

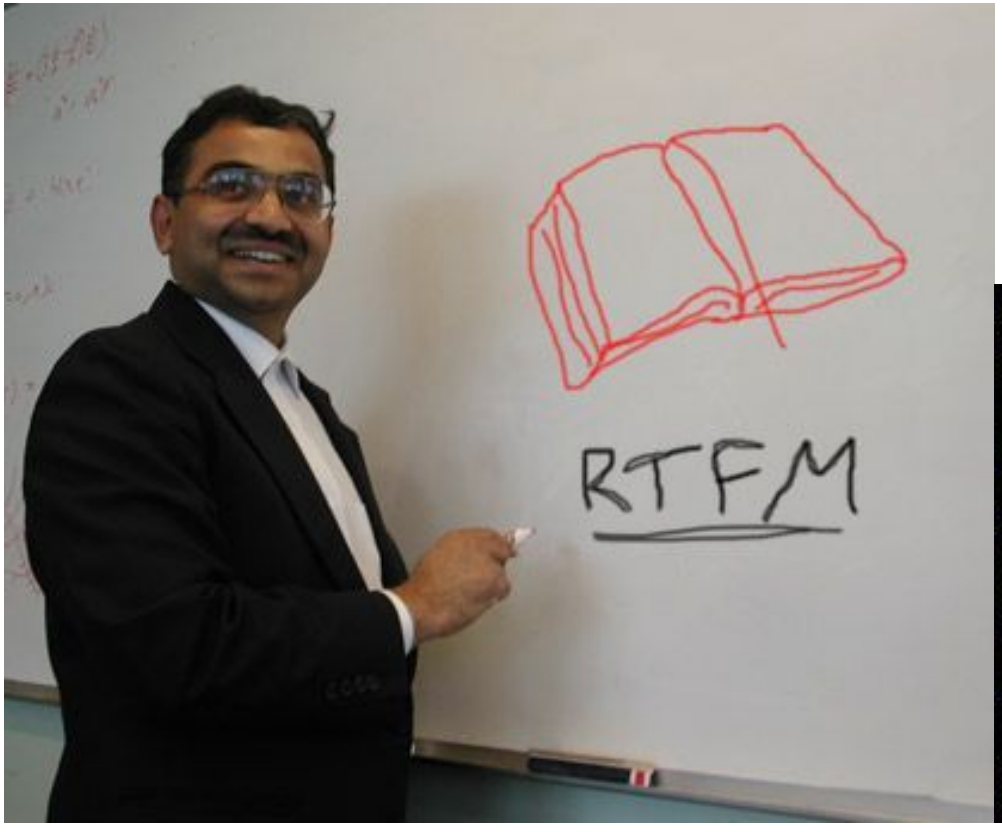
How to go from data to results?

How to go from data to results?

- Read the publication?

How to go from data to results?

- Read the publication?
- Read the manual?



How to go from data to results?

- 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.

Notebooks: interactive computational environments

Notebooks combine:

1. code fragments that can be executed,
2. text for the description of the application and
3. figures illustrating the data or the results.

In [2]:
```
import numpy as np
import matplotlib.pyplot as plt
```

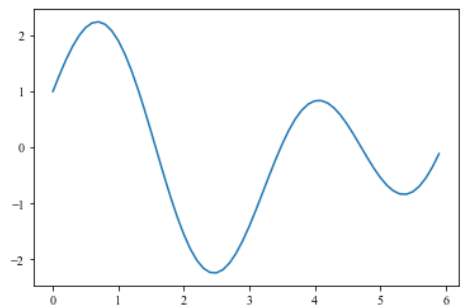
Data

Let's create a simple function.

In [6]:
```
x = np.arange(0, 6, .1)
y = np.cos(x) + 1.5 * np.sin(2 * x)
```

Make a simple plot

In [7]:
```
plt.plot(x, y)
plt.show()
```



Notebooks: interactive computational environments

Notebooks combine:

1. code fragments that can be executed,
2. text for the description of the application and
3. figures illustrating the data or the results.

"Digital Playground"

"Data Story Telling"

"Computational Narratives"

Notebooks: interactive computational environments

NATURE | TOOLBOX

Interactive notebooks: Sharing the code

The free IPython notebook makes data analysis easier to record, understand and reproduce.

Helen Shen

05 November 2014

http://www.nature.com/news/interactive-notebooks-sharing-the-code-1.16261

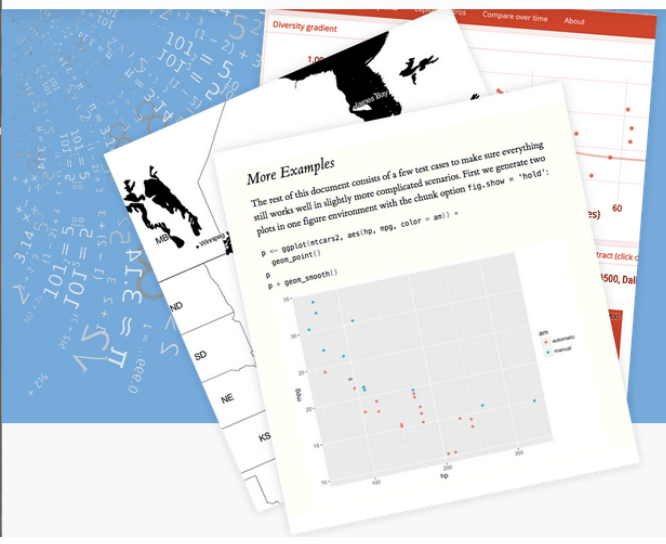
Interactive environments:

what exists today?

R Markdown

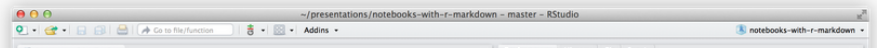
from RStudio

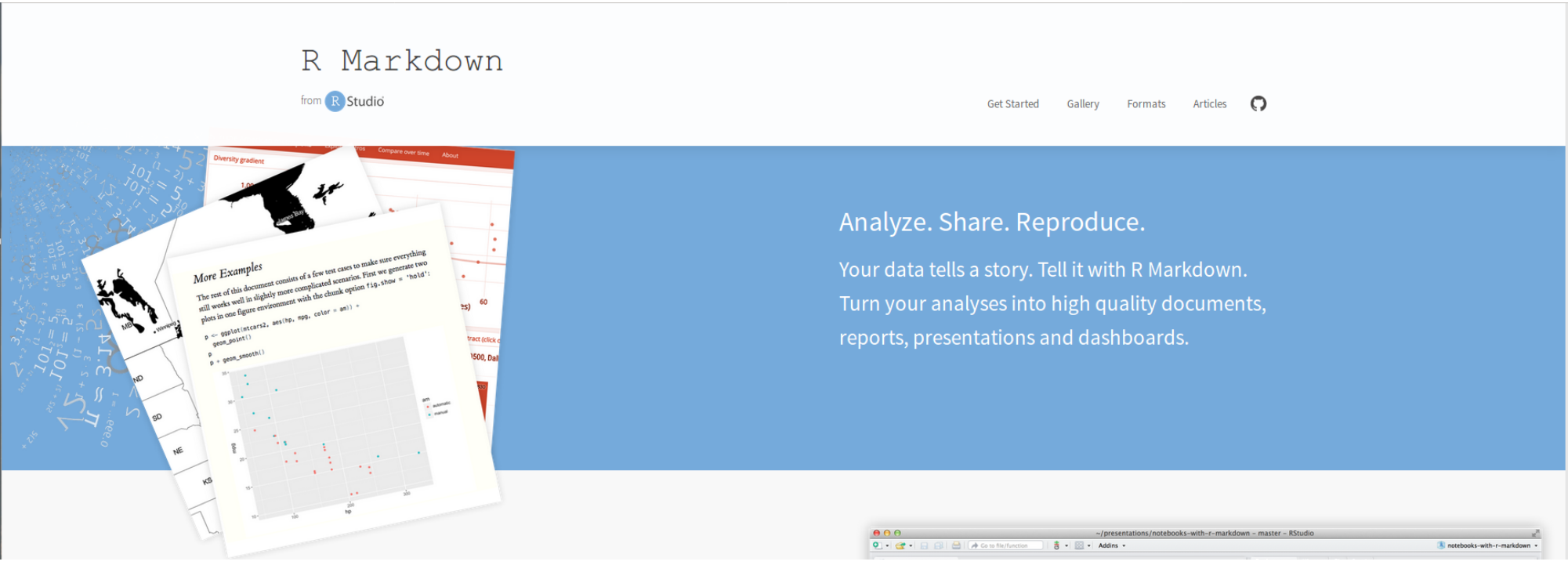
Get Started Gallery Formats Articles



Analyze. Share. Reproduce.

Your data tells a story. Tell it with R Markdown. Turn your analyses into high quality documents, reports, presentations and dashboards.





http://rmarkdown.rstudio.com/

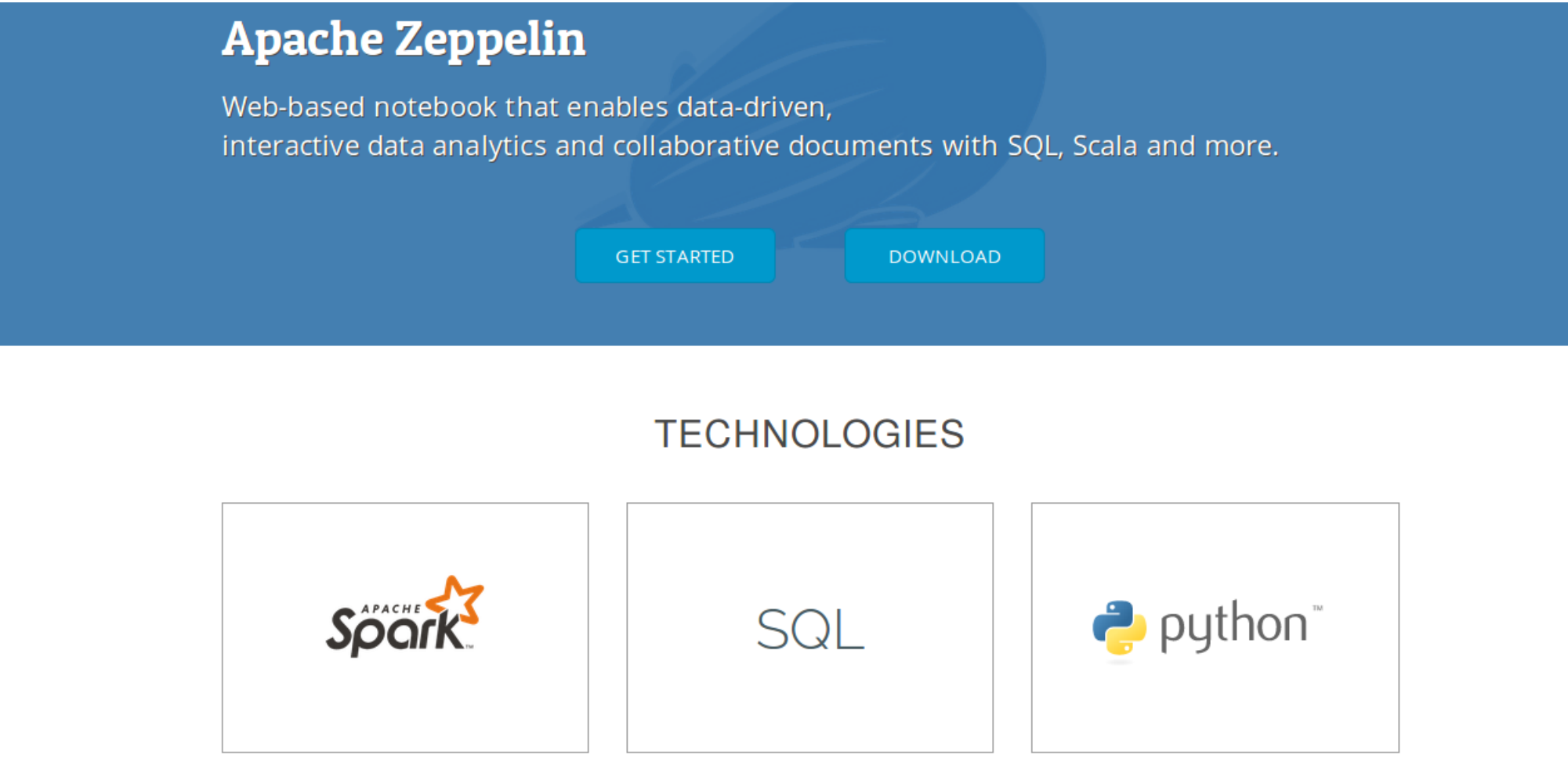
- Possible to export in journal or presentation formats
  https://github.com/rstudio/rticles
- LaTeX templates for different journals

Apache Zeppelin

Web-based notebook that enables data-driven,
interactive data analytics and collaborative documents with SQL, Scala and more.

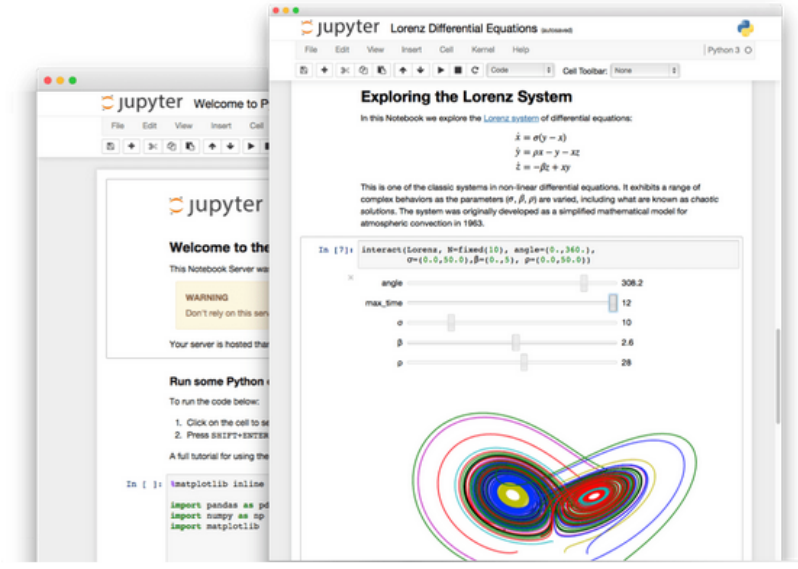
GET STARTED DOWNLOAD

TECHNOLOGIES



https://zeppelin.apache.org/

- 👍 Languages can be mixed in the same notebook
- 👍 Users can write their own interpreter (*language backend*)



The Jupyter Notebook

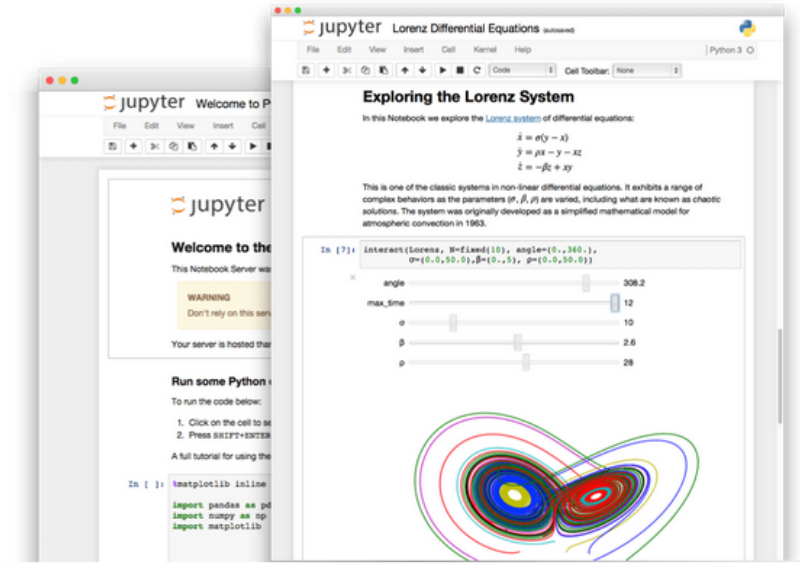
The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

Language of choice

Share notebooks

Interactive widgets

Big data integration



The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and much more.

</> Language of choice

Share notebooks

Interactive widgets

Spark Big data integration

`http://jupyter.org/`

- More than 40 language *kernels* available
- Can be used as a multi-user server (`jupyterhub`)

Follow Beaker

BEAKER

THE DATA SCIENTIST'S LABORATORY

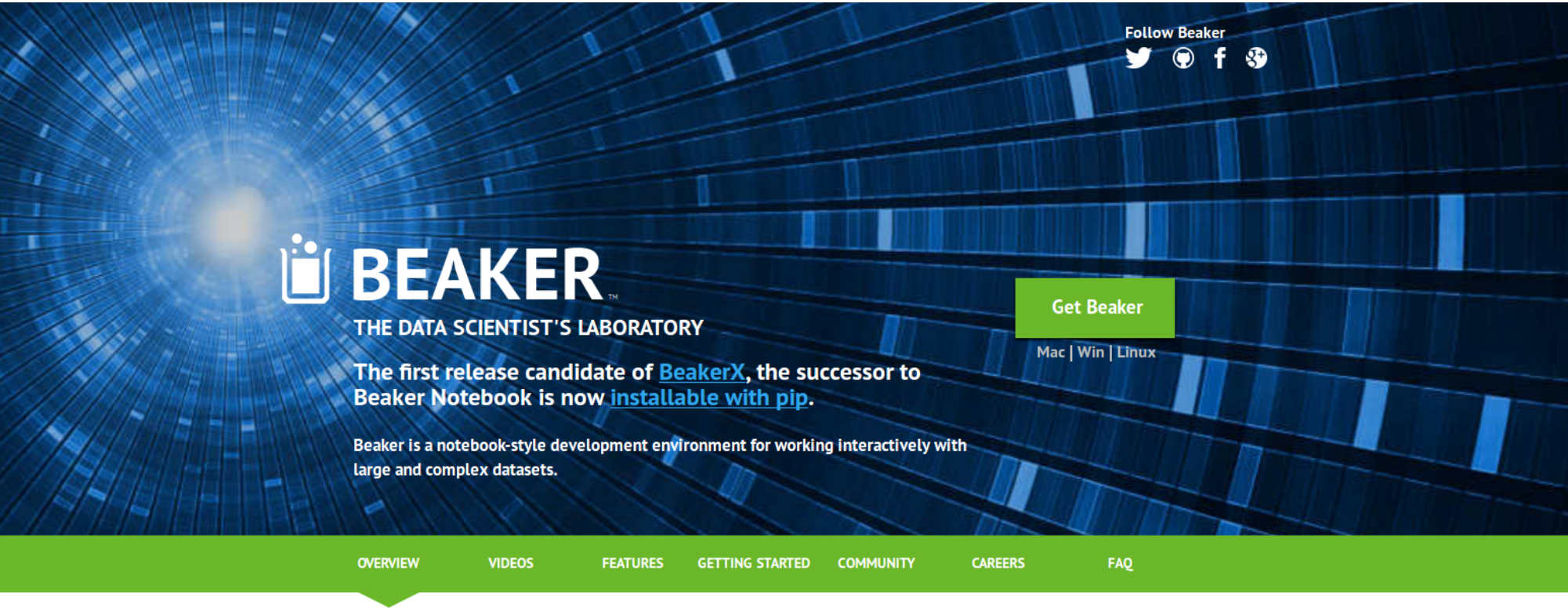
The first release candidate of BeakerX, the successor to Beaker Notebook is now installable with pip.

Beaker is a notebook-style development environment for working interactively with large and complex datasets.

Get Beaker

Mac | Win | Linux

The Perfect Tool for Iterative Exploration



The Perfect Tool for Iterative Exploration

`http://beakernotebook.com/`

- 👍 Usage of different languages in different cells, within the same notebook
- 👎 Installation and multi-language

COCALC

Collaborative Calculation in the Cloud

Run CoCalc

or sign in with your account

`https://cocalc.com/` "*Collaborative Calculation in the Cloud*"

- 👍 Support of many languages
- 👍 Users to upload their files on the platform

Exercise 1

subsetting using nco

Goal: extract a regional subset from field

Tool: `nco`

NetCDF subsetting

Notebook file: `NetCDF-regridding/netCDF-subsetting.ipynb`

Language: bash (not usual)

1. Run the notebook cell-by-cell ▶
2. Create a new notebook File / New Notebook
3. Modify the bounding box and perform subsetting
4. Repeat the operations on one of your files
5. Export the notebook as a pdf file

Exercise 2

regridding using nco

Goal: re-interpolate

Tool: `nco` + `ESMF`

NetCDF regridding

Notebook file: `NetCDF-regridding/netCDF-regrid.ipynb`

Language: bash

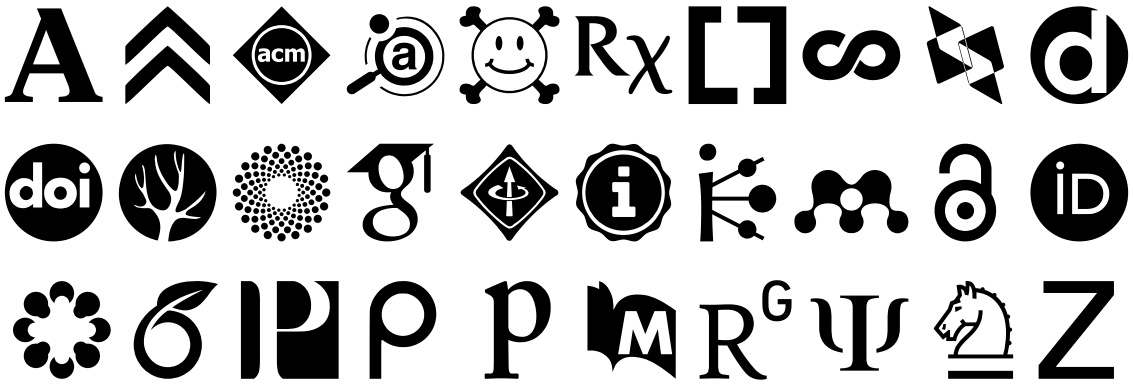
1. Run the notebook cell-by-cell ▶
2. Create a new grid file with a different resolution ℹ
3. Perform again the regridding
4. Try the regridding on one of your file

ℹ use `create-netCDF-grid.ipynb`
if you don't know how to create such a file

Another step

towards reproducibility

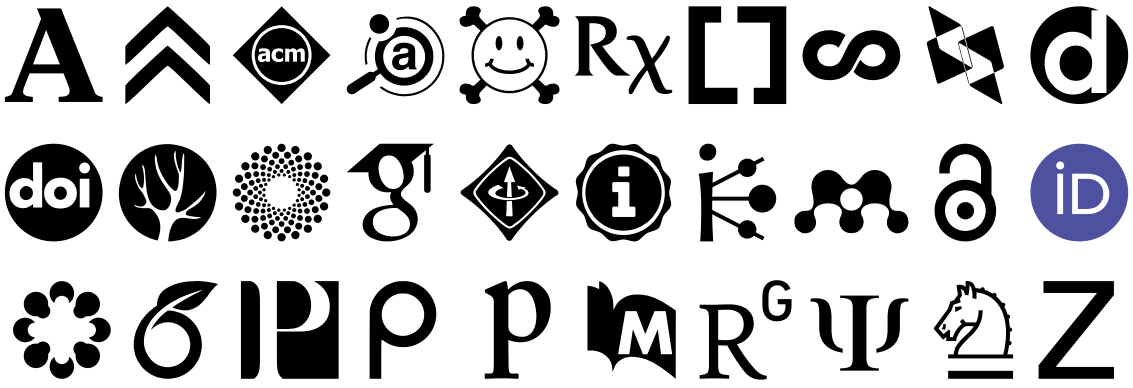
How to identify scientists/researchers?



Source: Academicons

How to identify scientists/researchers?

Let's work with ORCID



Source: Academicons

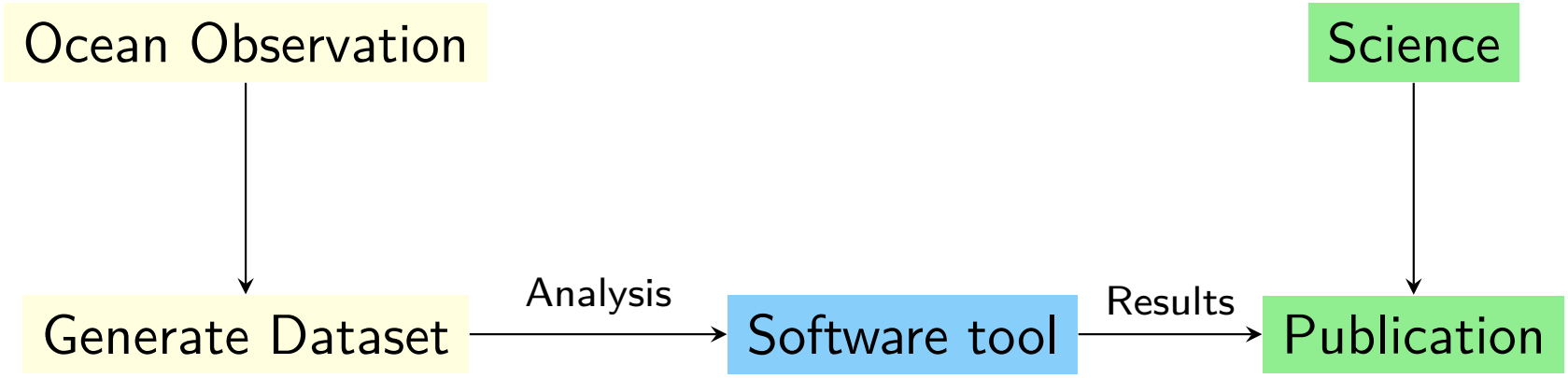
How to identify datasets and publications?

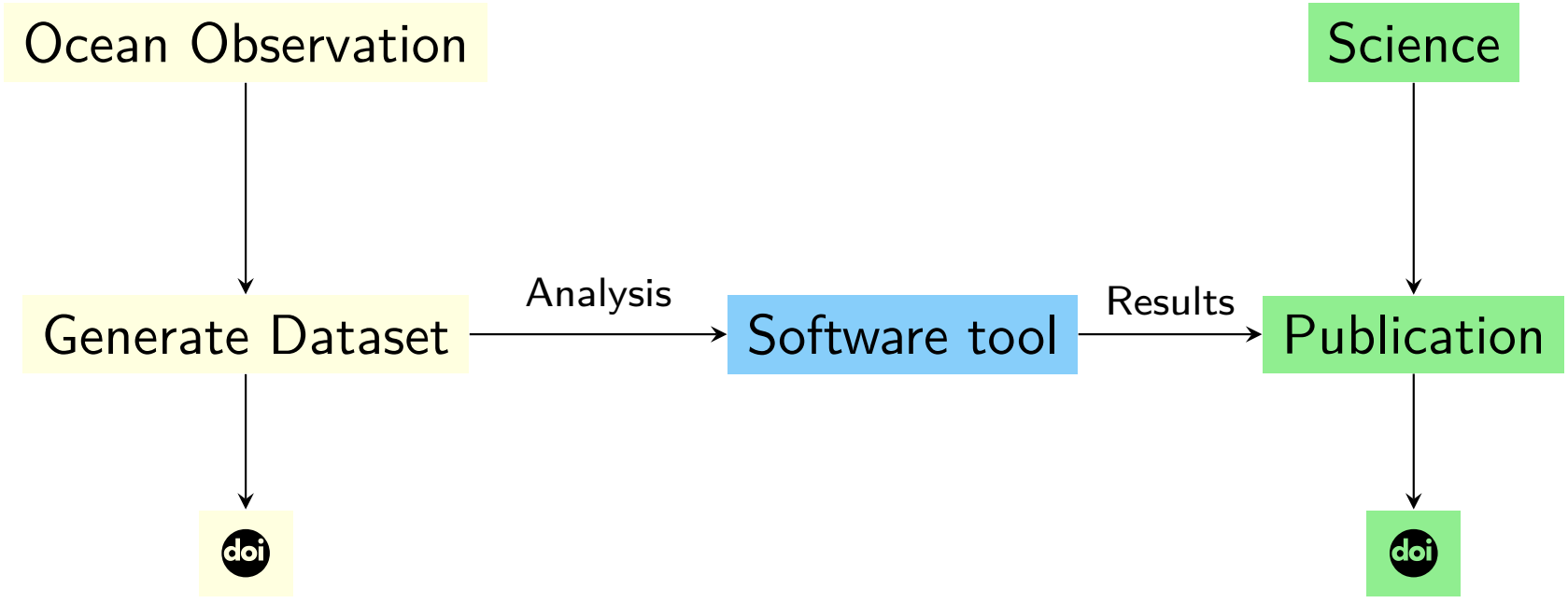
Digital object identifier

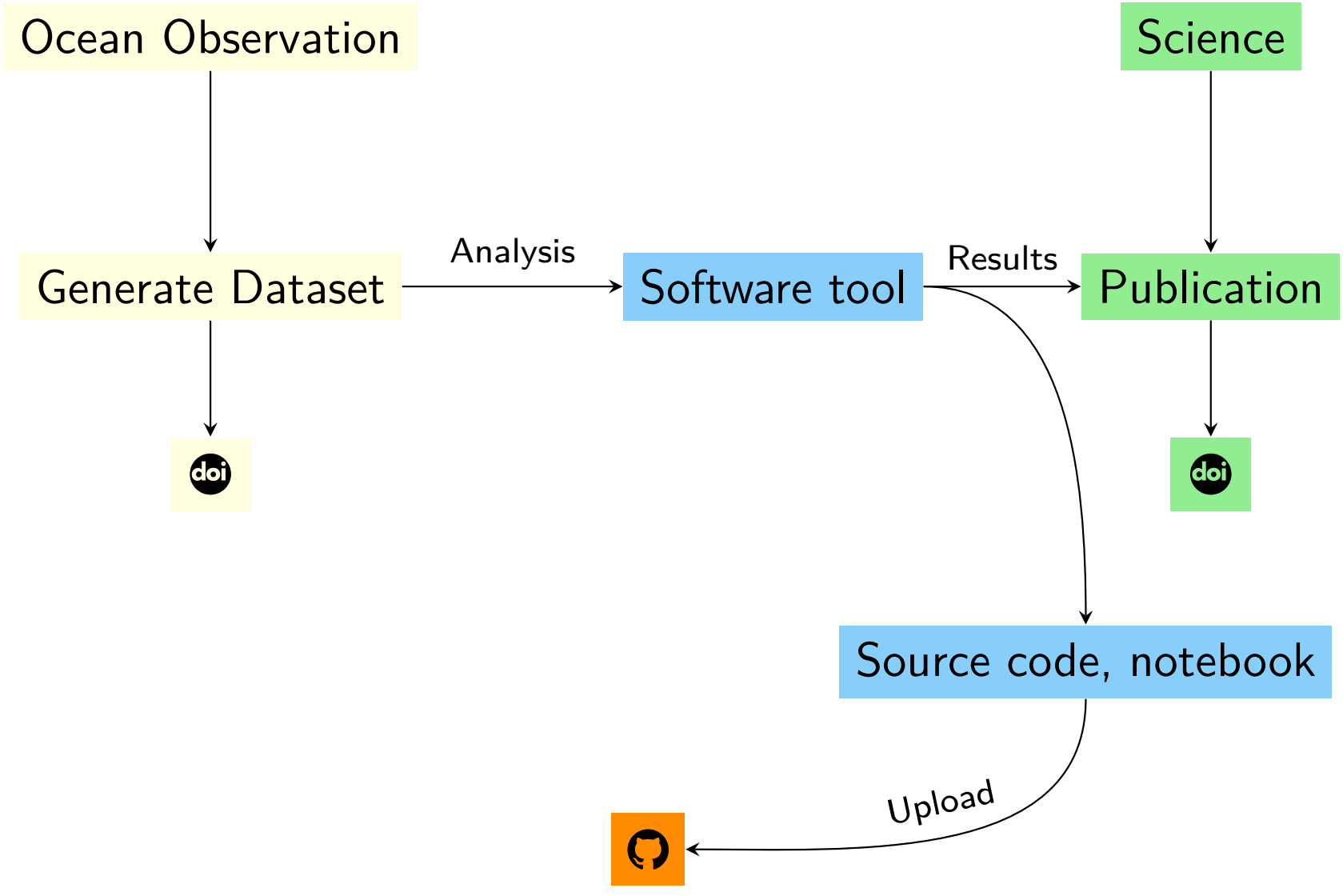
Ocean Observation

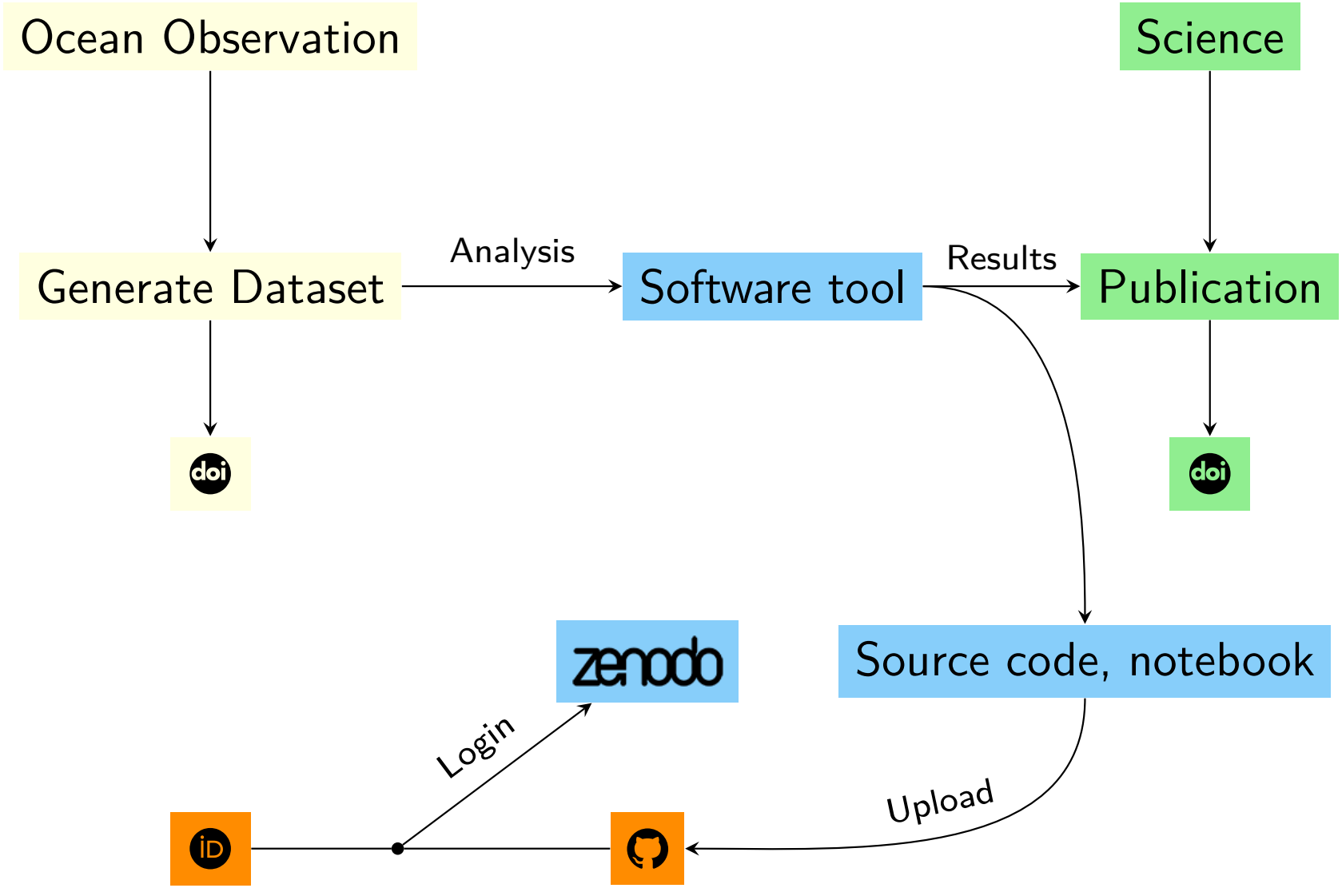
Science

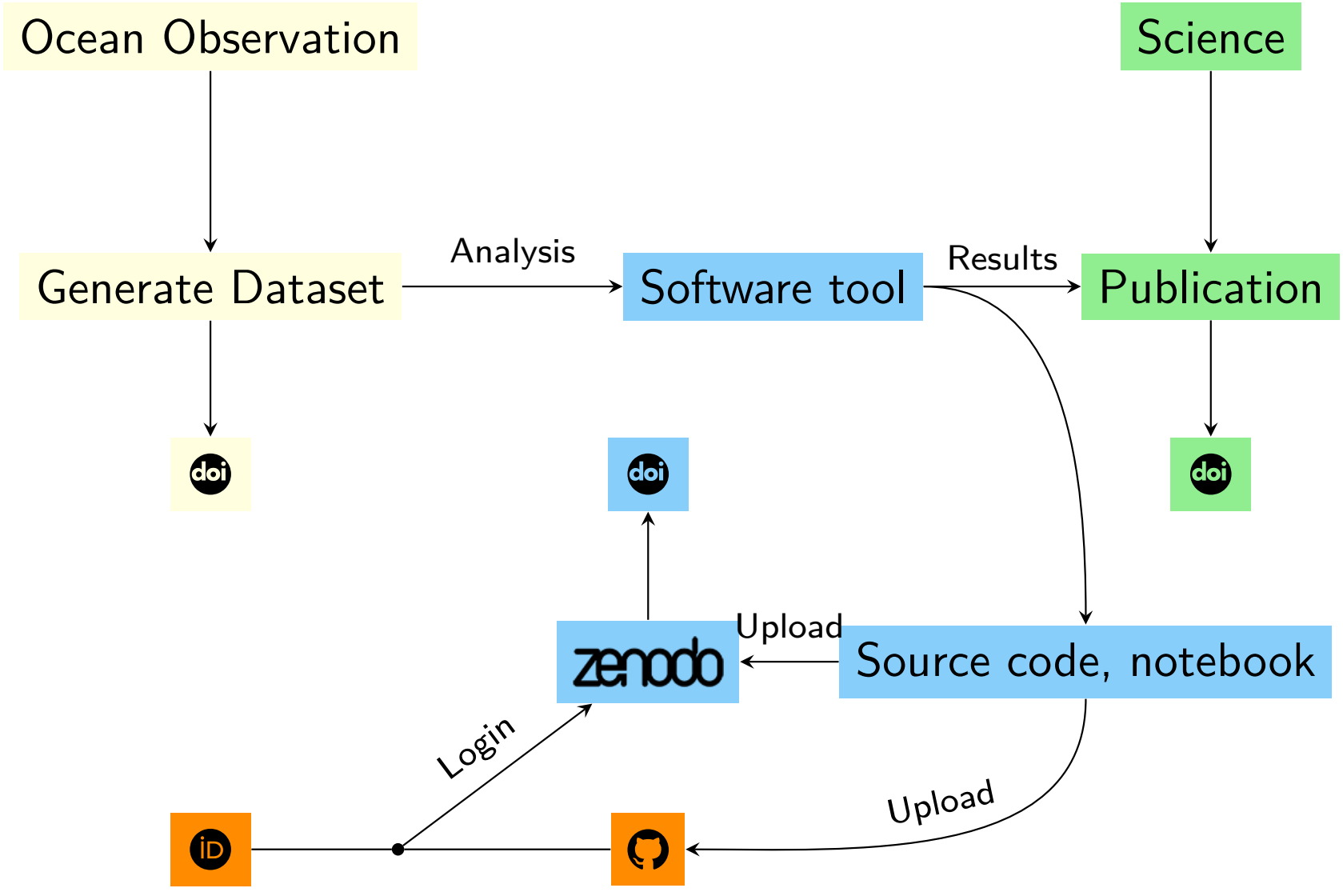

Ocean Observation
Generate Dataset
Science

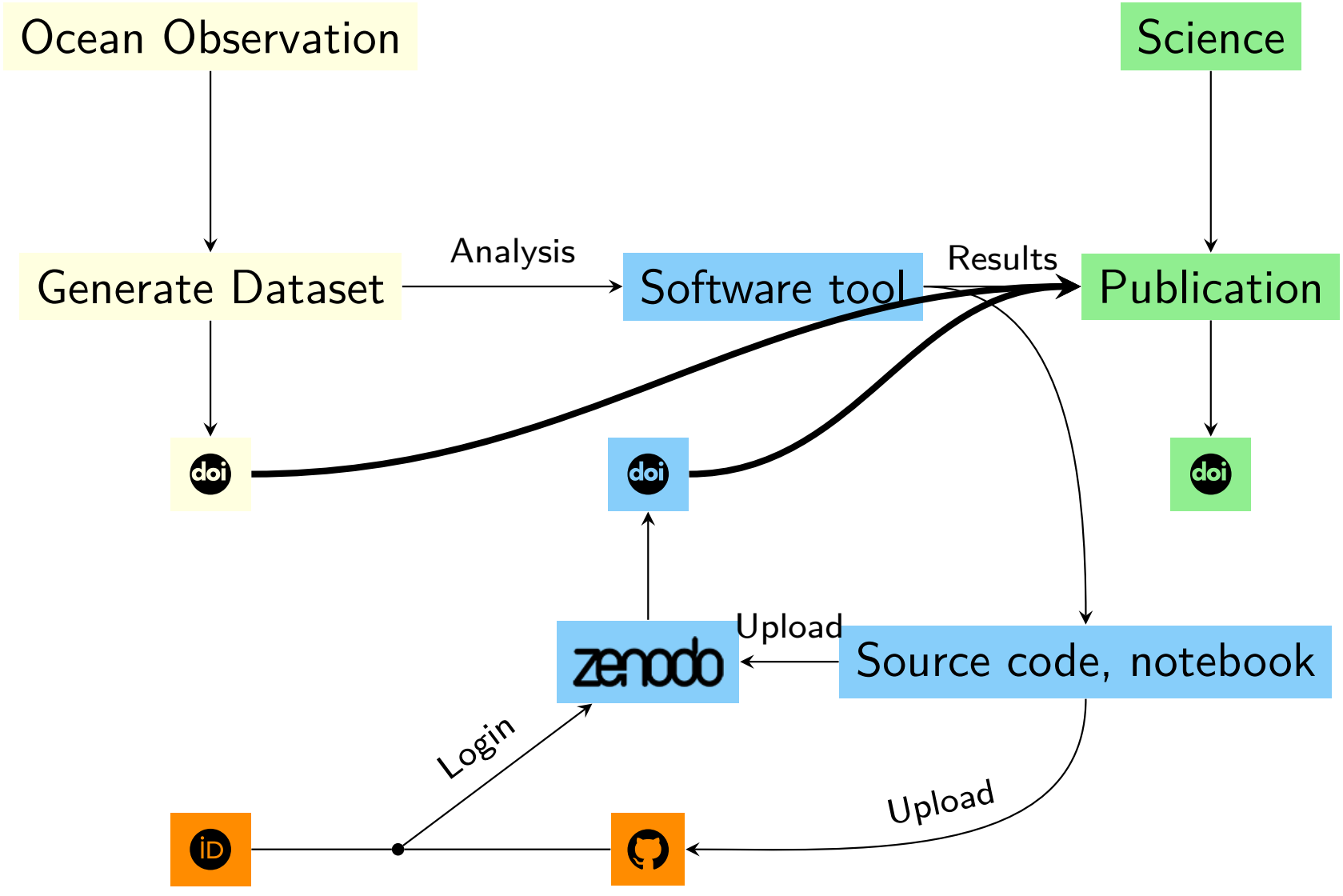

Ocean Observation
Science
Generate Dataset
Analysis
Software tool
Results
Publication


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


Ocean Observation
Science
Generate Dataset
Analysis
Software tool
Results
Publication
doi
doi
Source code, notebook
Upload


Ocean Observation
Science
Generate Dataset
Analysis
Software tool
Results
Publication
doi
doi
zenodo
Source code, notebook
Login
Upload


Ocean Observation
Science
Generate Dataset
Analysis
Software tool
Results
Publication
doi
doi
doi
zenodo
Upload
Source code, notebook
Login
Upload


Ocean Observation
Science
Generate Dataset
Analysis
Software tool
Results
Publication
doi
doi
doi
Upload
zenodo
Source code, notebook
Login
Upload

Zeno-what???

Zeno-what???

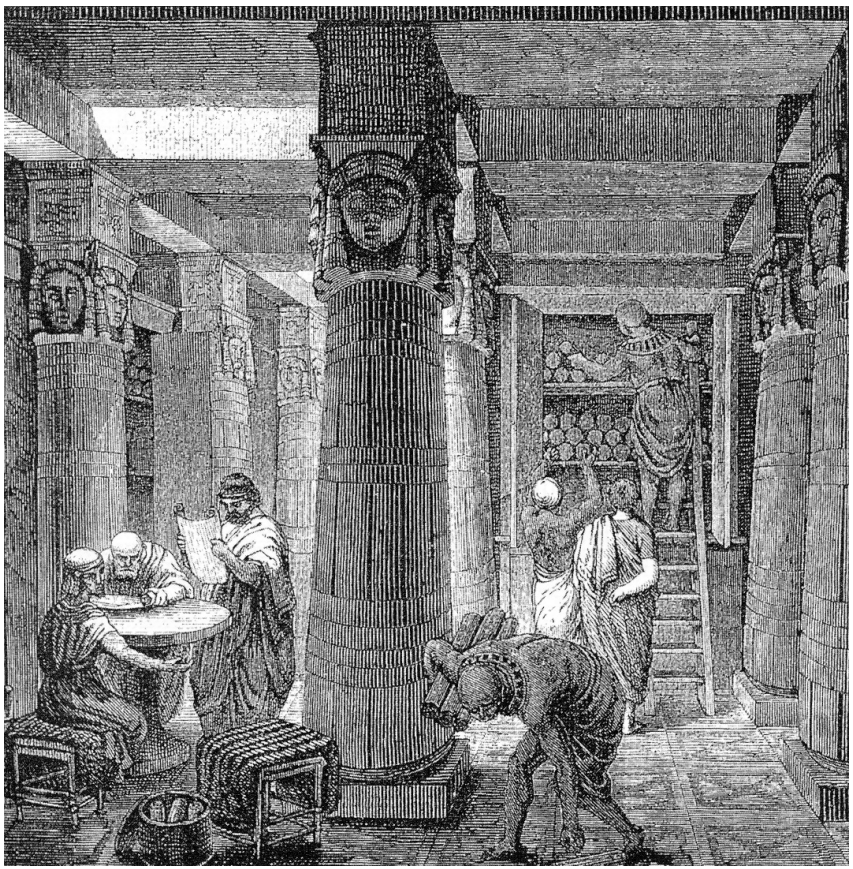
Zenodo – http://zenodo.org/

A platform to upload papers, datasets, software codes...
and to get permanent identifiers

Zeno-what???

After *Zenodotus*

1st superintendent of the Library of Alexandria and 1st critical editor of Homer



By O. Von Corven - Tolzmann, Don Heinrich, Alfred Hessel and Reuben Peiss. The Memory of Mankind. New Castle, DE: Oak Knoll Press, 2001, Public Domain,
https://commons.wikimedia.org/w/index.php?curid=2307486

Recent uploads

January 29, 2018 (v2.1.0) Dataset Open Access

View

Genome assemblies for "Versatile genome assembly evaluation with QUAST-LG"

Alla Mikheenko; Andrey Prjibelski; Vladislav Saveliev; Dmitry Antipov; Alexey Gurevich

De novo genome assemblies of Yeast_PB (S. cerevisiae, genome size: 12.1 Mb): Canu, FALCON, Flye, MaSuRCA (from Illumina pair-ends and PacBio SMRT) Yeast_NP (S. cerevisiae, genome size: 12.1 Mb): Canu, Flye, MaSuRCA (from Illumina pair-ends and Oxford Nanopores) Worm_PB (C. elegans,genome...

Uploaded on January 31, 2018

2 more version(s) exist for this record

April 14, 2016 (v2) Software Open Access

View

ambitcli-3.0.2

Jeliazkova, Nina; Kochev, Nikolay; Jeliazkov, Vedrin

A command line Java application used for processing chemical files, structure standardization, import into AMBIT database and processing AMBIT database entries. More information at http://ambit.sourceforge.net /download_ambitcli.html Chemical structure standardization option available since...

Uploaded on January 11, 2018

1 more version(s) exist for this record

January 9, 2018 (v3) Dataset Open Access

View

A global network of biomedical relationships derived from text

Percha, Bethany; Altman, Russ B.

Zenodo now supports DOI versioning!

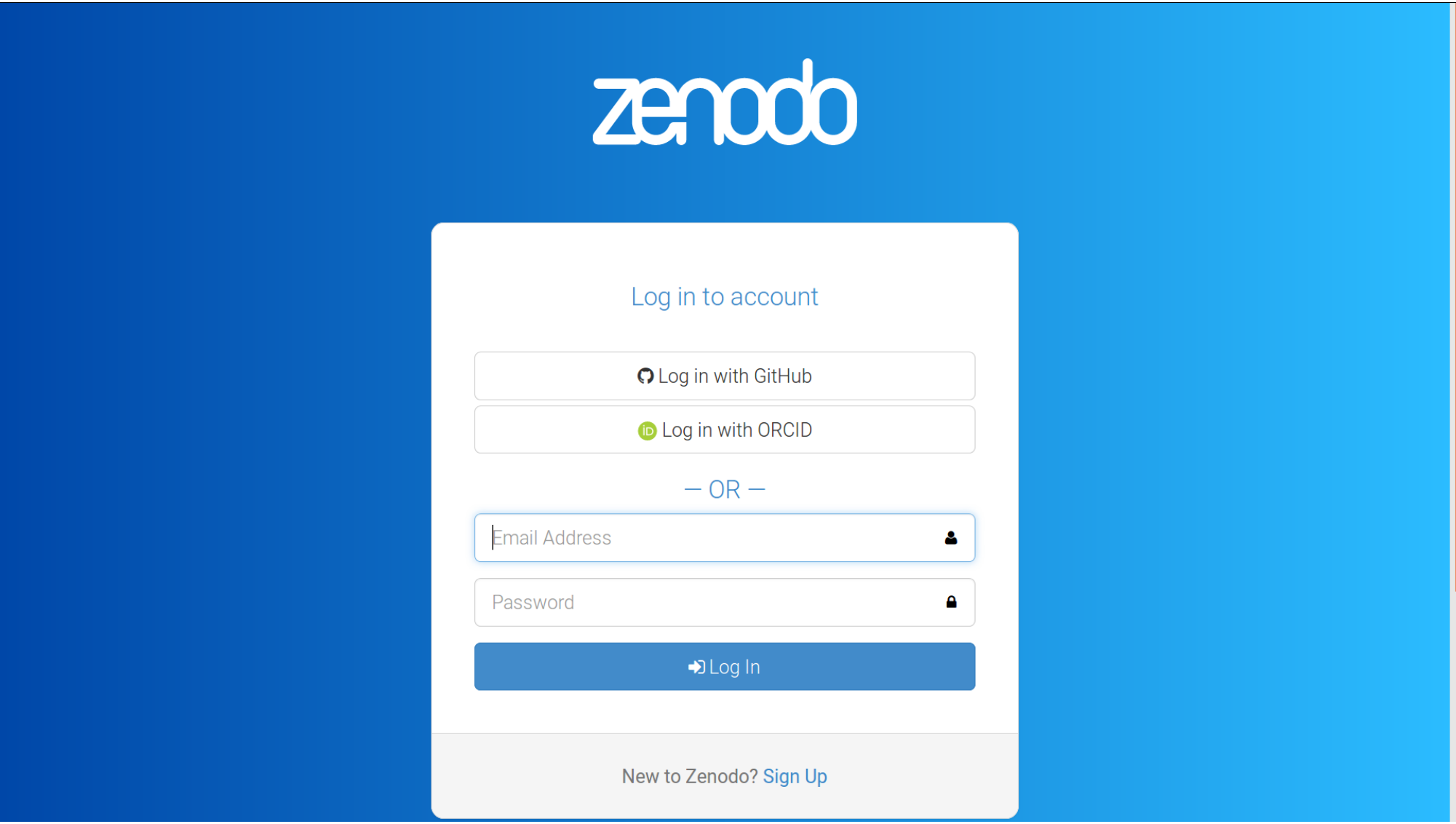
Read more about it, in our newest blog post.

Using GitHub?

Just Log in with your GitHub account and click here to start preserving your repositories.

Zenodo in a nutshell

- **Research. Shared.** — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
- **Citeable. Discoverable.** — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- **Communities** — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
- **Funding** — identify grants, integrated in reporting lines for research funded by the


zenodo
Log in to account
Log in with GitHub
Log in with ORCID
— OR —
Email Address
Password
Log In
New to Zenodo? Sign Up

Settings

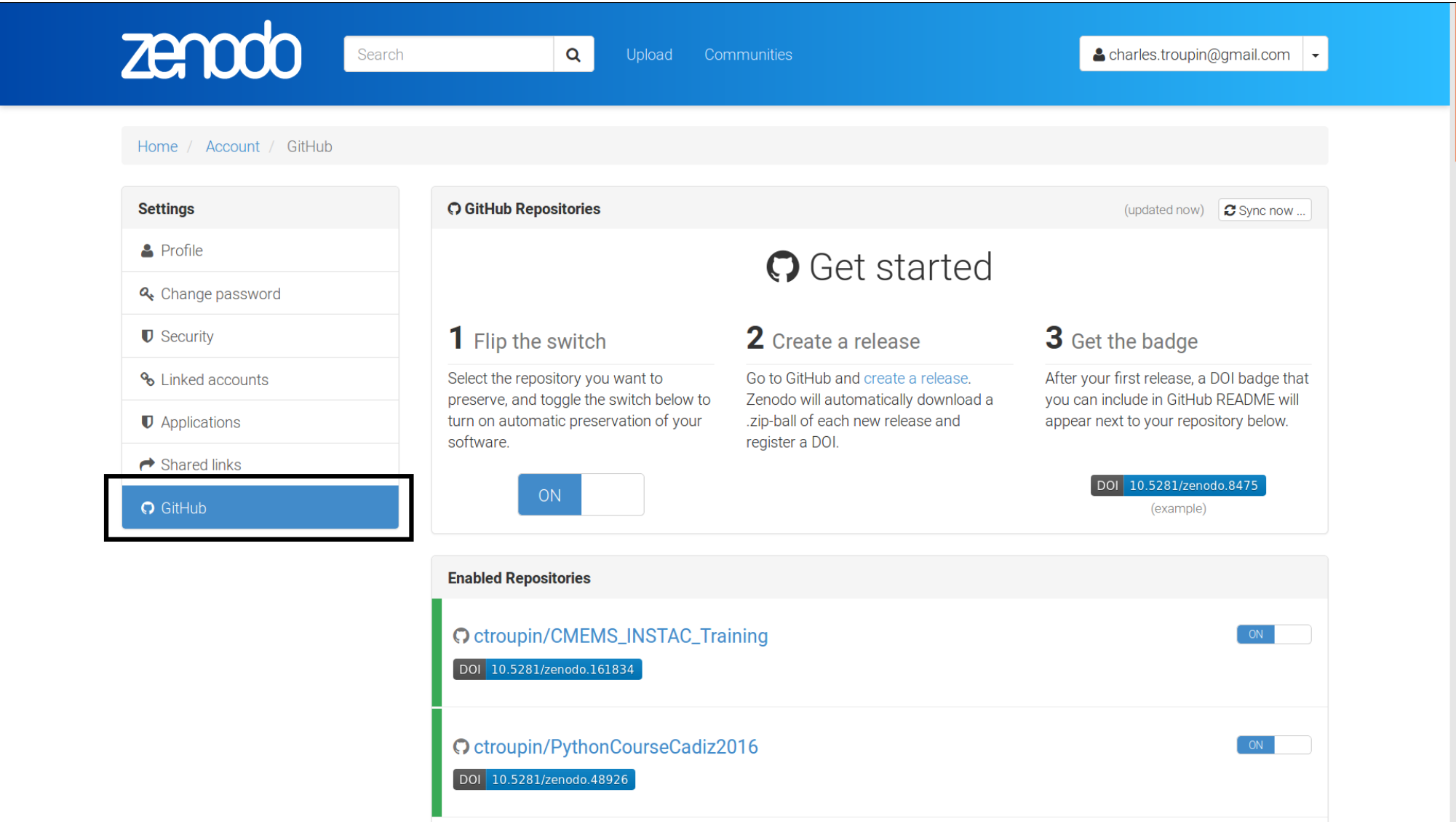
- Profile
- Change password
- Security
- Linked accounts
- Applications
- Shared links
- GitHub

Linked accounts

Tired of entering password for Zenodo every time you sign in? Set up single sign-on with one or more of the services below:

GitHub ✔ — Disconnect
Software collaboration platform, with one-click software preservation in Zenodo.

ORCID ✔ — Disconnect
Connecting Research and Researchers.


zenodo
Search
Upload
Communities
charles.troupin@gmail.com
Home / Account / GitHub
Settings
Profile
Change password
Security
Linked accounts
Applications
Shared links
GitHub
GitHub Repositories
(updated now)
Sync now ...
Get started
1 Flip the switch
Select the repository you want to preserve, and toggle the switch below to turn on automatic preservation of your software.
ON
2 Create a release
Go to GitHub and create a release. Zenodo will automatically download a .zip-ball of each new release and register a DOI.
3 Get the badge
After your first release, a DOI badge that you can include in GitHub README will appear next to your repository below.
DOI 10.5281/zenodo.8475
(example)
Enabled Repositories
ctroupin/CMEMS_INSTAC_Training
ON
DOI 10.5281/zenodo.161834
ctroupin/PythonCourseCadiz2016
ON
DOI 10.5281/zenodo.48926

July 31, 2017

Software Open Access

gher-ulg/DIVA: v4.7.1

Sylvain Watelet; Charles Troupin; Jean-Marie Beckers; Alexander Barth; Mohamed Ouberdous

New features

- Major feature : bottom analysis is now possible. The distance is counted from the bottom ocean, derived from the interpolation of the topography topo_fine.grd. This topography can be different (and finer) than topo.grd used for the creation of contours.
- Major feature : conversion of EMODnet bathymetry to Diva-readable format with the tool emobath2ghertopo.
- Major feature : Variable correlation length depending on the gradient of the depth. Advection field adapted to this relative length field. Suited for bottom analyses. Updated programs : diva3Ddat, divarlvargraddepth, rlvargraddepth.f90, divadoall, divaUVtopo, UVtopogen.f, divadocommit.
  - Acknowledgements field in 3D and 4D netCDFs.
  - New compilation option DIVAHUGEMEMORY. Enabled by default in divacompile_options, it allows the use of a finer mesh, and/or a larger domain. Particularly useful with variables characterized by low correlation lengths.

Bug fixes

- divacutNCDF : correction on climatology bounds + dealing with very big obsid vector
- divacompile_options : new tests on nc-config, due to recent change in its behaviour
- dv3Dreadnc.F : warning if dimensions are incoherent between GridInfo.dat and the netcdf file
- divadoall, divadoNCDF , divadoNCYR : corrected handling of 3DNCinfo and 3DNClist files so that recreating a new 4D NC file is much easier
- divadoxml : removed because deprecated (use divadoxml-gui instead)

Other

Update of the user guide

Preview

DIVA_v4.7.1.zip

Publication date:
July 31, 2017

DOI:
DOI 10.5281/zenodo.836727

Keyword(s):
Data-interpolation Ocean Sciences Oceanography Variational method SeaDataNet EMODnet

Related identifiers:
Supplement to:
https://github.com/gher-ulg/DIVA/tree/v4.7.1

Files (2.6 MB)

Name	Size	
gher-ulg/DIVA-v4.7.1.zip	2.6 MB	Preview Download

md5:cf9c3a19d2744a95f7bfa4a311b7729f

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.592476. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

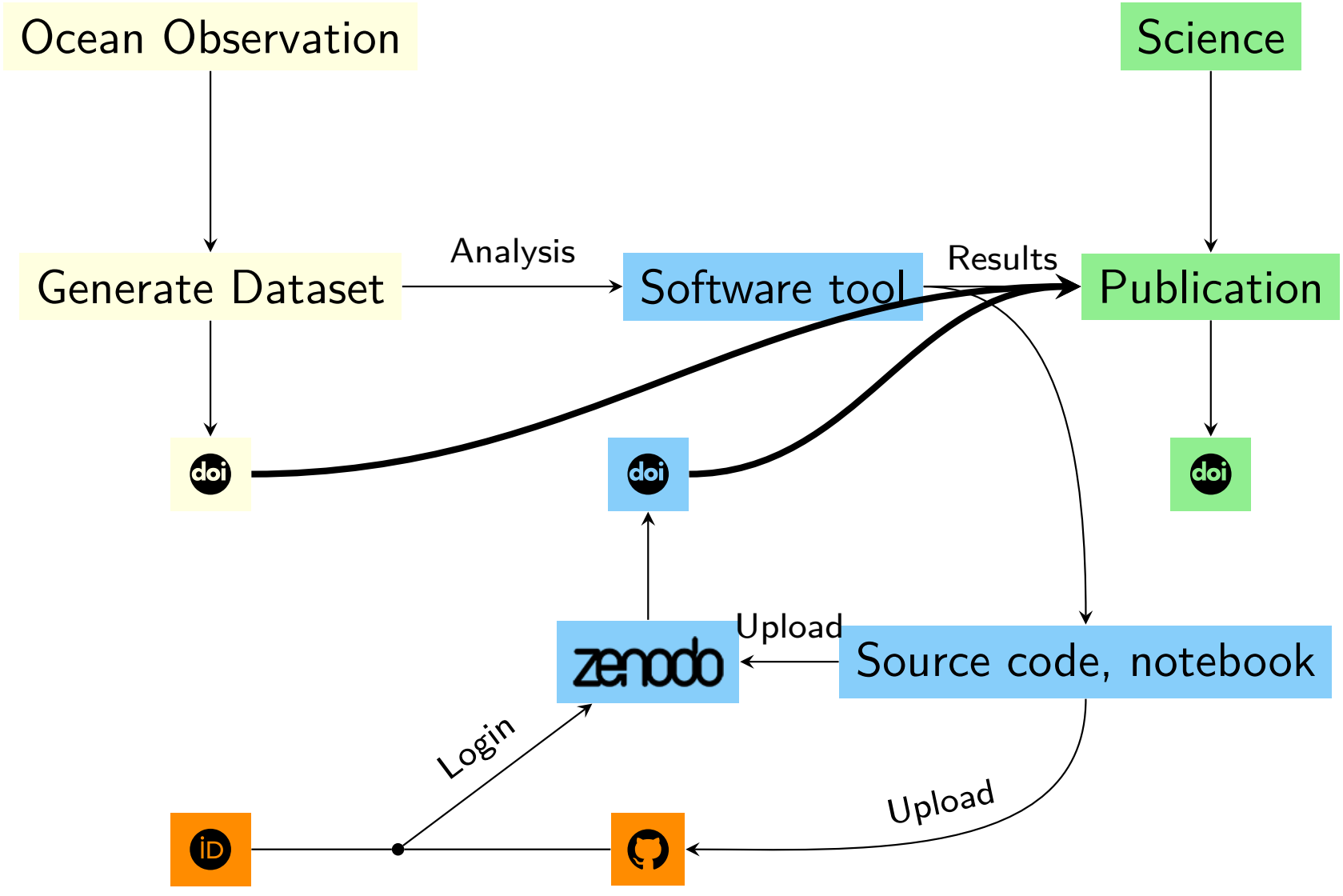
Cite as

Sylvain Watelet, Charles Troupin, Jean-Marie Beckers, Alexander Barth, & Mohamed Ouberdous. (2017, July 31). gher-ulg/DIVA: v4.7.1 (Version v4.7.1). Zenodo. http://doi.org/10.5281/zenodo.836727

Start typing a citation style...

Export

BibTeX CSL DataCite Dublin Core JSON JSON-LD MARCXML Mendeley


Ocean Observation
Science
Generate Dataset
Analysis
Software tool
Results
Publication
doi
doi
doi
zenodo
Upload
Source code, notebook
Login
Upload

A recent (and real) example: MedSea Atlas

Earth System Science Data
The Data Publishing Journal

https://doi.org/10.5194/essd-2018-9

Review article

02 Feb 2018

Mediterranean Sea Hydrographic Atlas: towards optimal data analysis by including time-dependent statistical parameters

Athanasia Iona[1,2], Athanasios Theodorou[2], Sylvain Watelet[3], Charles Troupin[3], and Jean-Marie Beckers[3]
[1]Hellenic Centre for Marine Research, Institute of Oceanography, Hellenic National Oceanographic Data Centre, 46,7 km Athens Sounio, Mavro Lithari P.O. BOX 712 19013 Anavissos, Attica, Greece
[2]University of Thessaly, Department of Ichthyology & Aquatic Environment, Laboratory of Oceanography, Fytoko Street, 38 445, Nea Ionia Magnesia, Greece
[3]University of Liège, GeoHydrodynamics and Environment Research, Quartier Agora, Allée du 6-Août, 17, Sart Tilman, 4000 Liège 1, Belgium

Review status
This discussion paper is a preprint. It is a manuscript under review for the journal Earth System Science Data (ESSD).

Received: 30 Jan 2018 – Accepted for review: 31 Jan 2018 – Discussion started: 02 Feb 2018

Abstract. The goal of the present work is to provide the scientific community with a high-resolution Atlas of temperature and salinity for the Mediterranean Sea based on the most recent datasets available and contribute to the studies of the long-term variability in the region. Data from the Pan-European Marine Data Infrastructure SeaDataNet were used, the most complete and, to our best knowledge, of best quality dataset for the Mediterranean Sea as of today. The dataset is based on in situ measurements acquired between 1900–2015. The Atlas consists of horizontal gridded fields produced by the Data Interpolating Variational Analysis, where unevenly spatial distributed measurements were interpolated onto a 1/8° x 1/8° regular grid on 31 depth levels. Seven different types of climatological fields were prepared with different temporal integration of observations. Monthly, seasonal and annual climatological fields have been calculated for all the available years, seasonal to annual climatologies for overlapping decades and specific periods. The seasonal and decadal time frames have been chosen in accordance with the regional variability and in coherence with atmospheric indices. The decadal and specific periods analysis was not extended to monthly resolution due to the lack of data, especially for the salinity. The Data Interpolating Variational Analysis software has been used in the Mediterranean Region for the SeaDataNet and its predecessor Medar/Medatlas Climatologies. In the present study, a more advanced optimization of the analysis parameters was performed in order to produce more detailed results. The Mediterranean Region past and present states have been extensively studied and documented in a series of publications. The purpose of this Atlas is to contribute to these climatological studies and get a better understanding of the variability on time scales from month to

A recent (and real) example: MedSea Atlas

The publication? "*Mediterranean Sea Hydrographic Atlas...*"

10.5194/essd-2018-9

A recent (and real) example: MedSea Atlas

The publication? "*Mediterranean Sea Hydrographic Atlas...*"

10.5194/essd-2018-9

The author? Sissy Iona (Hellenic Center for Marine Research)

0000-0001-6878-4671

A recent (and real) example: MedSea Atlas

The publication? "*Mediterranean Sea Hydrographic Atlas...*"

10.5194/essd-2018-9

The author? Sissy Iona (Hellenic Center for Marine Research)

0000-0001-6878-4671

The data? MedSea – T and S observation collection V2

10.12770/8c3bd19b-9687-429c-a232-48b10478581c

A recent (and real) example: MedSea Atlas

The publication? "*Mediterranean Sea Hydrographic Atlas...*"

10.5194/essd-2018-9

The author? Sissy Iona (Hellenic Center for Marine Research)

0000-0001-6878-4671

The data? MedSea – T and S observation collection V2

10.12770/8c3bd19b-9687-429c-a232-48b10478581c

The products? MedSea – T and S Annual Climatology

10.5281/zenodo.1146976 via zenodo

A recent (and real) example: MedSea Atlas

The publication? "*Mediterranean Sea Hydrographic Atlas...*"

doi 10.5194/essd-2018-9

The author? Sissy Iona (Hellenic Center for Marine Research)

iD 0000-0001-6878-4671

The data? MedSea – T and S observation collection V2

doi 10.12770/8c3bd19b-9687-429c-a232-48b10478581c

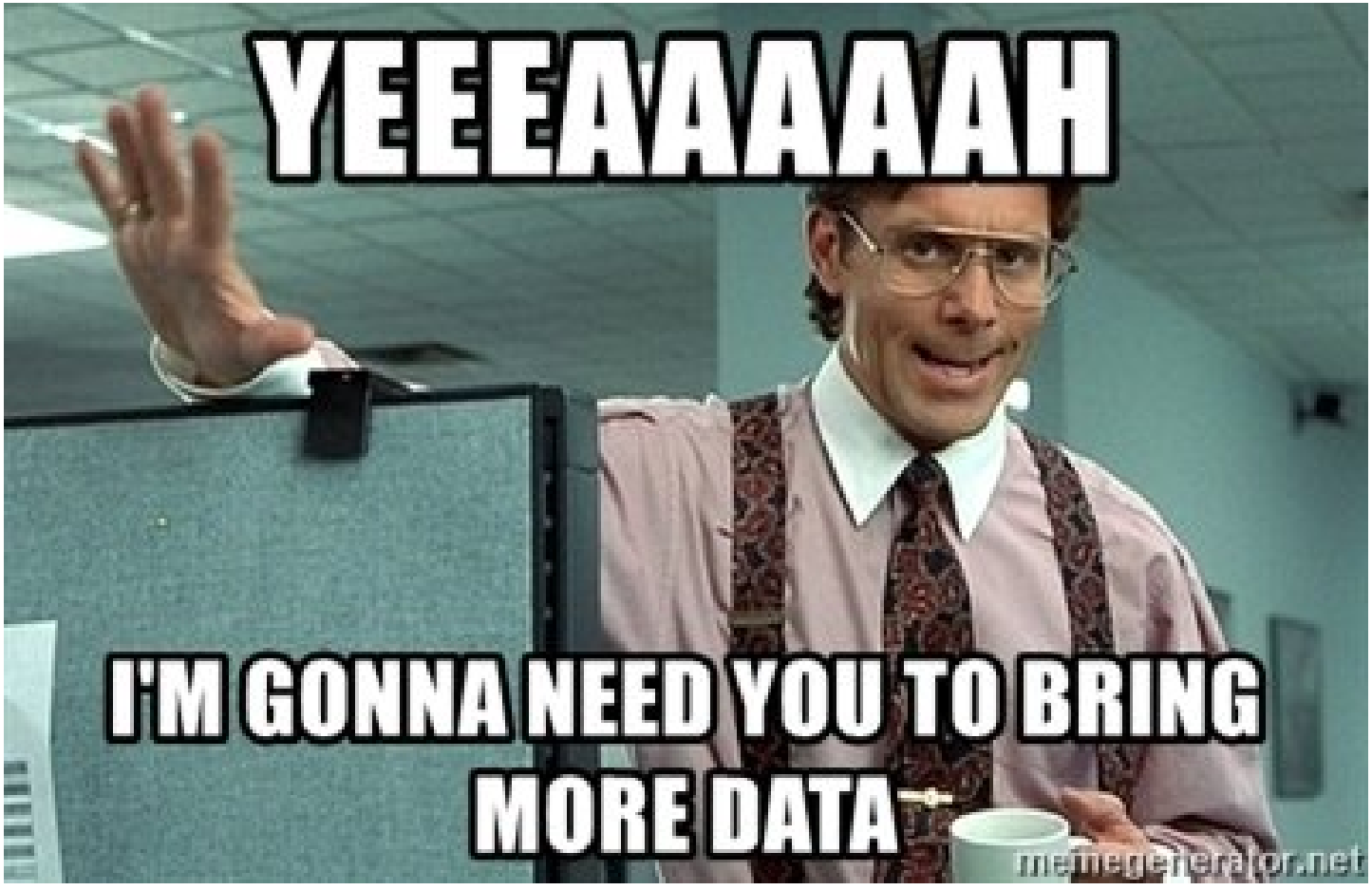
The products? MedSea – T and S Annual Climatology

doi 10.5281/zenodo.1146976 via zenodo

The method/tool? DIVA Version 4.6.11

doi 10.5281/zenodo.400970 via zenodo

~~Conclusions~~ Suggestions



~~Conclusions~~ Suggestions

1. Create an account in ORCID (https://orcid.org/)

2 minutes

~~Conclusions~~ Suggestions

1. Create an account in ORCID (https://orcid.org/) — 2 minutes
2. Create an account in Zenodo (http://zenodo.org/) — 2 minutes

~~Conclusions~~ Suggestions

1. Create an account in ORCID (https://orcid.org/) — 2 minutes
2. Create an account in Zenodo (http://zenodo.org/) — 2 minutes
3. Publish your code along with your paper — 15 minutes

~~Conclusions~~ Suggestions

1. Create an account in ORCID (https://orcid.org/) — 2 minutes
2. Create an account in Zenodo (http://zenodo.org/) — 2 minutes
3. Publish your code along with your paper — 15 minutes
4. Make your data public — ∞ minutes

Anything missing?

Working with

in situ observations

"*Without sufficient observations, useful prediction will likely never be possible*"

"*Models will evolve and improve, but, without data, will be untestable, and observations not taken today are lost forever.*"

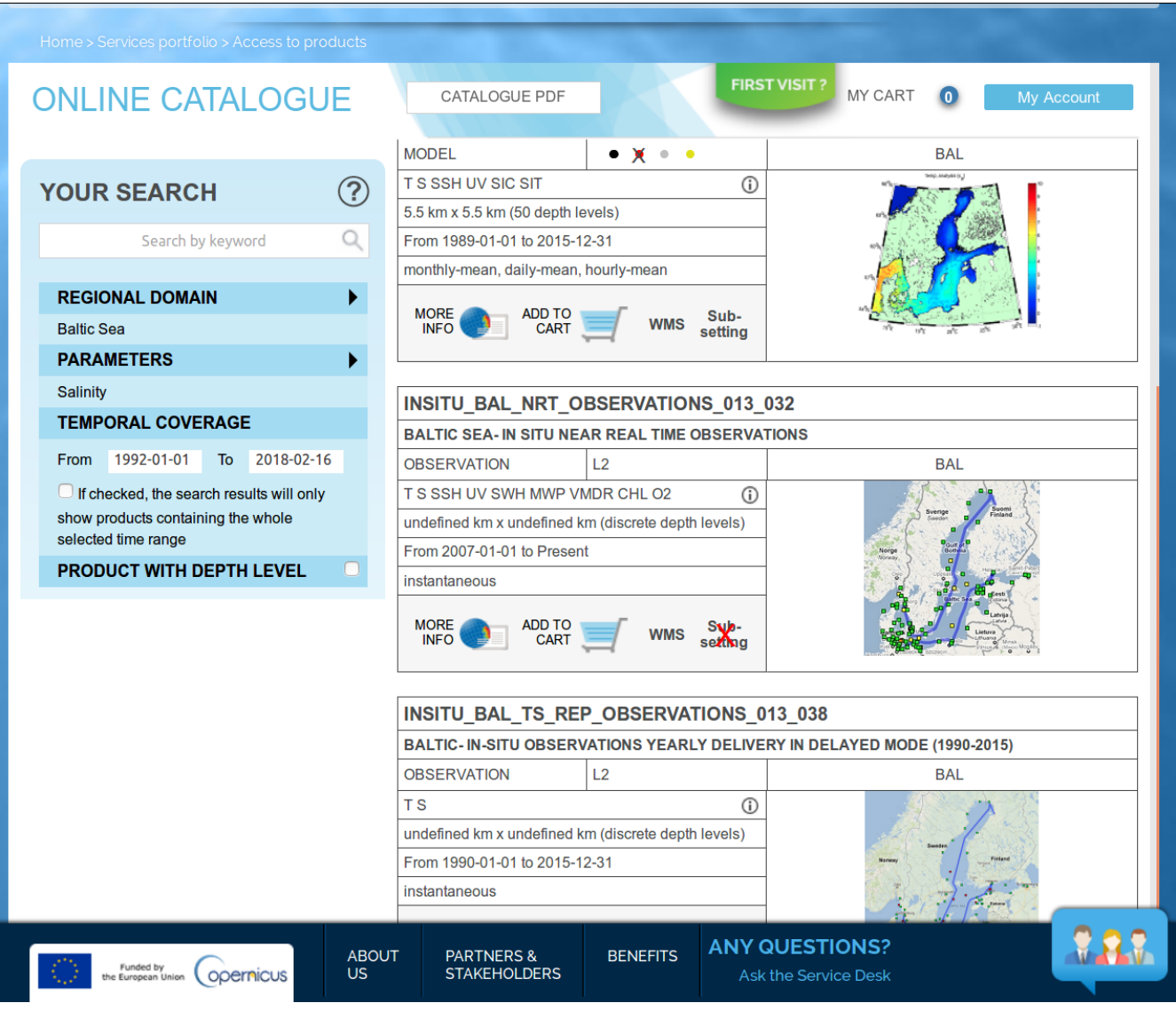
C. Wunsch et al. (2010) PNAS

"Without data assimilation, any attempt to produce reliable forecasts is almost certain to end in failure."

https://www.metoffice.gov.uk/learning/making-a-forecast/first-steps

How to access the data?

Copernicus Marine Environment Monitoring Service (CMEMS) catalog: http://marine.copernicus.eu



How to access the data?

1. Create user
2. Login into the system
3. Select the data set of interest from the catalog
4. Download from FTP

Exercise 3: find sea water temperature near Hamburg

using in situ data

In situ temperature

Data: folder `INSITU_BAL_NRT_OBSERVATIONS_013_032`
`index_latest.txt`: list of files with bounding box, time coverage etc `latest` directory: netCDF files

Tool: notebook `CMEMS_INSTAC/read_CMEMS_indexfile.ipynb`

Tasks:

1. Read the index file
2. Represent the data points on a figure (map)
3. Find the closest data point
4. Read the temperature at the point

Creating a presentation from your notebook

http://nbviewer.jupyter.org/github/gher-ulg/COST-EUMETSAT-Training/blob/master/CMEMS_INSTAC/slides_insitu.ipynb

How to access the data?

In Situ Thematic Assembly Center:

http://www.marineinsitu.eu/dashboard/